DOCUMENT RESUME

ED 336 457 UD 028 208

AUTHOR Rosser, Phyllis
TITLE Gender and Testing.
SPONS AGENCY National Commission on Testing and Public Policy.
PUB DATE Oct 89
NOTE 144p.
PUB TYPE Collected Works - General (020) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Blacks; Black Students; College Entrance Examinations; Comparative Testing; Females; High Schools; *High School Students; Males; Mathematics Tests; *Problem Solving; Questionnaires; Racial Bias; *Racial Differences; Sex Bias; *Sex Differences; *Standardized Tests; Suburban Schools; Test Bias; Testing Problems; Urban Schools; Verbal Tests; White Students
IDENTIFIERS Analogies; *Scholastic Aptitude Test

ABSTRACT

Test-taking differences between the sexes on the Scholastic Aptitude Test (SAT) were studied to determine if girls and boys, both white and African American, had different problem-solving styles that affected their ability to respond correctly on the test. Two studies are described. The first study, "An Investigation of the Performance Differences of White and Black Female and Male High School Students on Eight Verbal Questions from the November 1987 Scholastic Aptitude Test" (conducted by C. Ross), studied the performance of 10 white and 10 black high school juniors from two urban and two suburban high schools in the Seattle area, who were interviewed about their solution processes. There were no major sex differences in problem-solving styles, but blacks appeared to process analogies differently than did whites. The second study, "An Investigation of the Performance Differences of White and Black Female and Male High School Students on Five Math Questions from the November 1987 Scholastic Aptitude Test" (conducted by K. Kelly-Benjamin), studied solution methods used by 10 black female, 10 black male, 10 white female, and 10 white male high school juniors in New York City. In this sample, which performed well, females tended to rely more on mathematics knowledge than on test-taking skills. An analysis of the two studies, "Do Females and Males Have Different Problem Solving Styles that Create or Contribute to the Gender Gap" (P. Rosser), suggests that test-taking skills are crucial for SAT performance. A 38-item list of references is included. Thirteen appendices present supplemental data and SAT results. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Phyllis Rosser
Equality in Testing

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

GENDER AND TESTING

ED364457

OCTOBER 1989

Phyllis Rosser

Director, The Equality in Testing Project

Holmdel, New Jersey

This research was commissioned by the National Commission on Testing and Public Policy, Graduate School of Education, University of California, Berkeley.

UD028208

2 BEST COPY AVAILABLE

TABLE OF CONTENTS

I. An Overview of Gender Bias in Standardized Tests and Its Possible Causes . . . . . 1 - 11

II. Gender Differences in Problem Solving: An Introduction to the Research Studies. . . . . 12 - 13

III. An Investigation of the Performance Differences of White and Black Female and Male High School Juniors on Eight Verbal Questions from the November 1987 Scholastic Aptitude Test
A Research Study by Catherine Ross . . . . . 14 - 53

IV. An Investigation of the Performance Differences of White and Black Female and Male High School Students on SAT Math Questions
A Research Study by Kathleen Kelly-Benjamin. . . . . 54 - 93

V. Do Females and Males Have Different Problem Solving Styles That Create or Contribute to the SAT Gender Gap?
An Analysis with Policy Recommendations by Phyllis Rosser . . 94 - 102

VI. Bibliography and References . . . . . 103 - 105

VII. Appendices. . . . . 106 - 138

Appendix A: Profiles of Students Participating in Verbal Study. . . . 106
Appendix B: Questionnaire Given to Each Student in Verbal Study . . . 107
Appendix C: Student Response Form for Verbal Study. 108
Appendix D: Scoring Sheet for Verbal Study. 111
Appendix E: Five Problems Used in Math Study. 119
Appendix F: Process-Sequence Coding System for SAT Math Interviews. . 125
Appendix G: Interviewer Observation Forms for Math Study. 126
Appendix H: SAT Score Averages, 1967 - 1988 133
Appendix I: SAT Score Averages for the Class of 1988. 134
Appendix J: 1987-88 ACT Composite Scores. 135
Appendix K: Females' Average Scores are Lower Than Males' at Each Income Level. 136
Appendix L: Average SAT Scores for Males and Females on the November 1987 SAT by Grade Point Average (100,000 Sample) 137
Appendix M: November 1987 SAT Score Averages by Sex and Race (100,000 Sample). 138

AN OVERVIEW OF GENDER BIAS IN STANDARDIZED TESTS AND ITS POSSIBLE CAUSES

Where Gender Bias Occurs

Research on gender bias in testing began with the dawning of the second wave of feminism in the early 1970's. Educational Testing Service (ETS) Researcher Marlaine Lockheed (1973) identified three types of gender bias that frequently occurred on standardized achievement tests, reflecting societal differences in opportunities and expections for females and males. Known as "content bias" or "facial bias," these types included: quantity bias (the number of males and females referred to in a question), "stereotype bias" (the roles and activities in which each gender is portrayed); and "status bias" (the relative level at which each gender is presented). City University of New York Professor Carol Kehr Tittle investigated Lockheed's bias types on a large number of achievement tests (1974; 1978) and reported that many educational tests referred to males much more frequently than to females, showed men in higher status postions and depicted both sexes in stereotyped roles; women were housewives, teachers, nurses and secretaries while men were doctors, lawyers and politicians. Tittle stated that even if these depictions of a male-oriented world did not have a negative effect on girls' test scores, they were offensive in their perpetuation of cultural bias against females and should be eliminated.

ETS Researchers Ekstrom, Lockheed and Donlon (1979) analyzed major achievement test batteries for students in grades 1 though 12 and found that girls were more likely to succeed on a question when the people depicted were either female or "neutral" in sex rather than male. However, men continue to outnumber women in many tests (Selkow, 1984), including the Scholastic Aptitude Test (SAT) which is widely used for college admissions decisions and

scholarship awards (Rosser, 1987). The SAT also continues to refer to many more men than women. Most of the men are famous, the women usually are not (Rosser, 1987b).

Sex differences in socialization also present equity problems in the area of context validity, creating "context bias." Researchers have found that males and females perform differently on questions set in an experiential context that is familiar to them (Coffman, 1961; McCarthy, 1975; Rosser, 1989) For example, Rosser found questions about sports and finance were much more difficult for girls on the November 1987 SAT, while they outscored boys on several questions in a human relations context. Girls have traditionally preferred questions set in a humanities or aesthetic/philosopical context while boys prefer questions set in a science or business context (Donlon and Angoff, 1971).

Researchers have also found sex differences in skill areas with girls outperforming boys in computation (Doolittle, 1985) and algebra (Donlon, 1971) while boys scored higher in geometry and word problems (Doolittle, 1985; McPeek and Wild, 1987) Context and skill differences have to be carefully balanced on standardized tests in order to achieve an unbiased evaluation of the abilities of both sexes. In fact, the Stanford-Binet I.Q. test is rebalanced periodically to ensure that both sexes score equally well (Dwyer, 1979).

The tests where sex bias appears to have the greatest impact on females' educational opportunities are the two most widely-used college entrance examinations, the SAT published by ETS and the ACT Assessment published by American College Testing Program. These tests are biased in their predictive validity. The main purpose of these tests is to predict first year college performance but they do this less successfully for young women than for young

men. Females receive higher average grades than males in all subjects in high school and college (Cordes, 1986; Clark and Grandy, 1984) and lower average scores on both tests. In other words, the academic abilities of high school girls are systematically underpredicted while the abilities of high school boys are overpredicted.

The Effects of Biased Admissions Tests

Validity bias results in use bias. Inaccurate predictive validity means that college entrance examinations limit young women's access to competitive colleges and high school talent search programs, keep them from receiving millions of dollars in merit scholarship awards, and cause them to downgrade their college aspirations. By underpredicting their academic performance, the tests affect women's chances to gain entrance to over 1,500 four year colleges and universities that require SAT or ACT scores or use cut-off scores for admission (Rosser, 1987). Lower scores also create lowered expectations, causing women to apply to less competitive colleges and universities than their grades would indicate. A 1987 Carnegie Foundation report found that 62 percent of the students questioned had lowered their college expectations after receiving their SAT or ACT scores (Boyer, 1987).

Lower SAT and ACT scores also exclude girls from Talent Search programs which offer summer enrichment and accelerated courses to seventh through eleventh graders with high SAT (Verbal-430+; Math-500+) or ACT scores. These programs, originally developed by Johns Hopkins' Center for the Advancement of Academically Talented Youth (CTY), are offered to thousands of students at major state and private universities around the country (Rosser, 1989).

Women also lose millions of dollars in merit scholarship awards as a result of lower SAT, ACT and Preliminary Scholastic Aptitude Test (PSAT) scores. The prestigious National Merit Scholarship Corporation which

distributes $23 million annually, selects its semi-finalist pool solely on the basis of PSAT scores. In 1987-88, women's PSAT averages were 44 points lower than men's so they received only 36 percent of the approximately 6,000 scholarships (National Merit Scholarship Annual Report, 1988). Four states - Massachusetts, Missouri, South Dakota, and Virginia - also award merit scholarships solely on the basis of SAT, ACT or PSAT scores with the result that boys are more likely to receive them (Rosser, 1989). When New York State attempted to join their ranks in 1989, Federal District Judge John M. Walker ruled that the use of SAT scores as the sole basis for awarding merit scholarships discriminates against girls and required that grades be used along with test scores on a 50/50 basis (Rosser, 1989). This ruling in a class action suit brought by the Women's Rights Project of the American Civil Liberties Union against the New York State Education Department, changed the criteria for awarding $8.2 million annually in the largest state merit scholarship program in the country.

All these factors contribute to a real dollar loss for women in later life as they get less prestigious jobs, earn less money and have fewer leadership positions. Of course, the life-long loss of self confidence cannot be measured in financial terms.

The Scholastic Aptitude Test

The SAT is taken annually by 1.5 million students - 52 percent of whom are female - so academic potential is underpredicted for 780,000 students. It is composed of two sections, Verbal and Math, each scored on a 200 to 800 point scale, with a maximum combined score of 1600. In 1988, girls' average SAT scores were 56 points lower than boys' - 13 points in the Verbal Section, an area where girls excelled until 1972 - and 43 points in the Math Section, where

girls have always scored lower than boys (Dwyer, 1976a).

ETS national score data collected since 1967 document young women's lower SAT averages for the past twenty-two years. Girls have consistently averaged 40 to 50 points lower than boys in the Math Section since national data were first collected (see Appendix H). It is particularly surprising that this gap has not narrowed during the past ten years as females' enrollment in math and science classes (Burton et al., 1988) and their career aspirations have risen dramatically. Girls averaged five to seven points higher than boys in the Verbal Section until 1972 when boys began outscoring them verbally, as well. This gap has gradually widened to 13 points.

ETS has been aware of the descrepencies between male and female performance over the years. Beginning in the 1950's, content specifications required that equal numbers of scientific and practical affairs questions (which boys prefer) and human relations and aesthetic-humanities questions (which girls prefer) be included in the Verbal test. Prior to this, test questions had been drawn from human relations and the humanities (Dwyer, 1976a). ETS researchers Donlon and Angoff (1971) reported that "balancing the content...contributes to a better balance for the scores between the sexes" (p. 25) However, ETS researcher Carol Dwyer (1976a) noted that no attempt was made to balance the math content where boys had always excelled "by a great number of points since the first administrations of the test" (p. 756)

Women of color are doubly penalized by the SAT. Women and men in every racial/ethnic group receive lower average composite scores than white men. However, women in every ethnic group, receive lower average scores than the men in their ethnic group. For example, African American men averaged 209 points lower than white men in 1988 but African American women averaged 32 points lower than African American men and 241 points lower than white men (see

Appendix I).

The Preliminary Scholastic Aptitude Test

A similar pattern of sex bias can be found on the Preliminary Scholastic Aptitude Test/National Merit Qualifying Test (PSAT/NMQT) taken by 1.1 million students (54 percent are female, so performance is underpredicted for 594,000 students). This test also contains a verbal and math section, each scored on a scale of 20 to 80 points (ETS claims that an approximation of future SAT scores can be obtained by multiplying scores by 10). In 1987, girls averaged 41 points lower than boys on the Math and 13 points lower on the Verbal (in SAT terms).

To qualify as a National Merit Scholarship Semifinalist, verbal scores are doubled and the math score is added in order to "give girls a better chance" (Rosser, 1987) However, as their verbal scores decline, they are not able to overcome the large math gender gap. As a result, only 34% of the Semifinalist pool in recent years has been female. National Merit Scholarship winners must replicate their PSAT scores with high SAT scores, which also works against girls (Rosser, 1989).

The ACT Assessment

Women's first year college performance is also underpredicted by the ACT Assessment taken annually by nearly a million students, primarly in the Midwest, Southwest and South. Fifty-four percent of the test takers are female so college grades are being underpredicted for nearly 540,000 students. The ACT is considered an achievement test rather an aptitude test, surveying acquired knowledge in four subject areas: English Usage, Mathematics Usage, Social Studies and Natural Science. Each section is scored on a scale ranging from 1 to 36, with men receiving higher average scores than women in all subject areas except English Usage.

In 1987-88, the average ACT Composite Score (the average of the combined subject scores) for men was 19.9 compared to 18.6 for women (ACT, 1987). However, Gamache and Novick (1985) found that all ACT subject scores and the ACT Composite score "consistently underpredicted" women's two year cumulative college Grade Point Average, even when partially controlling for different courses taken. Male and Female students of color also receive lower ACT Composite scores than white males and women within each ethnic group receive lower scores than men, as they do on the SAT (Appendix J).

Although the ACT plays a major role in determining college entrance, it has not been studied as extensively as the SAT because it is not preferred by the elite colleges and universities and is not as influential in determining scholarship winners. In January 1989, American College Testing Program announced that the test has been redesigned to emphasize a wider range of mathematical knowledge, more abstract reading skills and a new testing of scientific concepts. Since the first redesigned ACT will be administered in October 1989, no analysis of the test has been included in this study.

Identifying the Causes of the SAT Gender Gap

In October 1987, Phyllis Rosser received a Women's Educational Equity Act Program Grant from the Federal Department of Education to investigate the possible causes of women's lower SAT performance than men in every racial/ethnic group. This allowed her to conduct two item analyses (examine the responses of males and females to each question) to determine whether individual questions were creating the gender gap. An initial study was done on one form of the June 1986 SAT. Working with University of Vermont Sociology Professor James W. Loewen and Princeton Review President John Katzman, Rosser analyzed the responses of 1,112 Princeton Review coaching students from New York City, correlating their answers to self-reported grades,

test anxiety, perception of ability and parental background data obtained from a student questionnaire. Rosser then analyzed the responses of 100,000 students nationwide who took one form of the November 1987 SAT, correlating it with Student Descriptive Questionnaire information. All the data for this study were obtained from a College Board magnetic tape supplied by ETS.

These analyses identified a number of verbal and math questions which favored males by a wide margin (10 percent more boys than girls responded correctly) but only a few - on the verbal test - which favored females. Loewen/Rosser/Katzman found 13 questions on the June 1986 test where boys performed significantly better and 4 questions where girls were the better performers. Rosser found 21 questions on the November 1987 SAT which favored boys by a wide margin and only 3 which favored girls. All the math questions with large gender differences (10 on the June 1986 SAT and 17 on the November 1987 SAT) favored boys. It was concluded that the SAT was not being "balanced" properly. Too many questions favoring boys were allowed to remain on the test. (ETS tries out every question on the experimental section, so that the test writers know how the sexes will perform.)

Verbal questions with large gender differences on both tests reflected cultural differences. Questions about relationships and appearance favored females while questions about science, business and traditionally male activities such as sports and war, favored males. It was found that the content of these questions, on both SATs, could be easily changed without affecting test validity, making the results more accurate for women.

The disparity in math questions was more complex and appeared to be caused by differences in test taking strategies. Girls on the November 1987 SAT moved through the math test more slowly than boys, answering fewer questions at the ends of the two math sections. A larger percentage of girls than boys also

omitted answers on 50 of the 60 math questions. This suggested that they were more reluctant to risk a guess if they didn't know the answer.

Both item analyses confirmed that SAT scores were underpredicting girls' high school performance and overpredicting boys.' On the June 1986 test, girls had better grades in math and English but equal scores on the Verbal test and lower scores on the Math test. On the November 1987 test, girls' grades were higher in both subject areas but they scored lower in both the Verbal and Math tests. Rosser also found that young women with the best high school grades (A to A+ GPA's) were penalized the most on the November 1987 SAT with a 23 point gender gap in the Verbal test and a 60 point gap in the Math. Since these were the women applying to the most competitive colleges and competing for merit scholarships, the test bias seemed worst for the brightest (see Appendix L).

Rosser and her associates expected that lower SAT score averages would have a negative impact on girls' self images and they did. In the June 1986 Study, both sexes were asked to estimate their verbal and math ability. Unfortunately these evaluations turned out to be more in line with their SAT scores than their high school grades in English and math - which was good for boys but bad for girls. Since high self confidence is important for a speeded test like this, girls appear to be handicapped by their lower self perceptions.

Rosser's November 1987 analysis included the first identification by an independent study (not conducted by ETS), of questions that created the gender gap for women in each racial/ethnic group. Questions were also identified that created large performance differences between women of color and white women but African American, Hispanic and Native American women scored worse on such a large number of questions that it was impossible to find a general pattern. However, Asian American women averaged better Math scores than all other women, including white women. In fact, their Math score averages were just 14 points

lower than white males (but 40 points lower than Asian males), while white females averaged 34 points lower than white males (see Appendix M).

Rosser's research also provided evidence that disputed several College Board assertions about the predictive validity of the SAT. Her findings challenge the College Board statement that more women than men taking the test are from low income, less educated families and that these women pull down the female score average. Although it has long been known that SAT scores rise as family income and parental education increase (Nairn, 1980), she found that young women at every income and parental educational level scored lower than the men from the same background (see Appendix K). In other words, ALL women were bringing the score averages down.

Rosser also questioned the often repeated College Board claim that young women take fewer years of high school math. She found their own data (1988 National Ethnic/Sex Profiles)) show that women now take as much algebra and nearly as much geometry as men - all the math that is required to succeed on this test. She reports that a longitudinal study of 4,000 students in Montgomery County, Maryland (Gross, 1988), found girls in the same advanced math classrooms as boys, taking the SAT on the same day, received higher average grades but lower average SAT scores by 37-47 points.

The College Board also attributes women's higher grades in college to the fact that they take "easy" humanities courses rather than engineering, math and science, where grading is more difficult (Rigol, 1989). However, Rosser found that the College Board's own validity studies show that women receive better college grades than men in all courses including math, science and engineering, despite their lower SAT scores (Clark and Grandy, 1984). Recent research by Lubetkin at Princeton (1988) and Kanarek at Rutgers (1988) has found that first year women at these universities had lower average SAT scores but higher

average first year grades in all subjects, including math and science.

Rosser notes that while other research shows a narrowing of cognitive differences between the sexes (Linn and Hyde, 1988; Fennema and Hyde, 1989; Feingold, 1988), the SAT gender gap is widening, particularly in the Verbal test. She also points out that a number of tests showing similar gender differences which are used by ETS to justify the SAT gap (such as those used for the NAEP and the High School and Beyond studies) are also written by ETS. She questions the success of ETS' review process and wonders why their tests show such consistently similar patterns of gender differences not found on other tests.

Rosser's WEEA Progam Grant research findings were published in April 1989, in The SAT Gender Gap: Identifying the Causes, which is available from the Center for Women Policy Studies, 2000 P Street, N.W., Washington, D.C. 20036.

GENDER DIFFERENCES IN PROBLEM SOLVING

An Introduction to the Research Studies

Do Females and Males Use Different Problem-Solving Styles?

The purpose of the two studies described in the following pages was to investigate further the test taking differences between the sexes on the SAT to see whether girls and boys, both white and African American, had different problem solving styles that affected their ability to respond correctly on the test. This also allowed further exploration of the possibility suggested by Rosser's November 1989 SAT findings, that females' test taking skills rather than cognitive differences or deficiencies were lowering their SAT Math scores.

Eight Verbal questions and five Math questions were selected for these studies: these were the questions which created the largest gender differences for students with the highest GPA's (A to A+ and B+ to A-) on the November 1987 SAT. The selected questions were given to junior high school students with higher than average PSAT scores and B+ or better GPA's. Twenty students participated in the Verbal study and forty students participated in the Math study. Half of each group was female and half male; half of each gender group was African American, the other half white.

An interviewer observed as the students solved the problems, noting cognitive and test taking differences, timing, and affective comments. Afterwards, each student discussed her/his problem solving approach and perception of the difficulty of each question. Unlimited time was provided for each question in order to evaluate how speededness effects responses for the different groups.

Although the sample sizes were not large enough to provide statistically

significant findings, it was hoped that they would indicate important areas for future research. Catherine Ross conducted the Verbal study with 20 public high school students from Seattle. Kathleen Kelly-Benjamin conducted the Math study with 40 students attending public, private and parochial high schools in New York City. Their findings are presented on the following pages, along with an overall analysis and comparison. Each study ends with suggestions for future research, including the necessity of replicating this work with larger samples.

AN INVESTIGATION OF THE PERFORMANCE DIFFERENCES OF

WHITE AND BLACK FEMALE AND MALE HIGH SCHOOL STUDENTS

ON EIGHT VERBAL QUESTIONS FROM THE NOVEMBER 1987 SCHOLASTIC APTITUDE TEST

The purpose of this exploratory study was to investigate the ways in which students process and respond to three different types of Scholastic Aptitude Test verbal questions: sentence completions, analogies, and antonyms. The study investigated the response processes utilized by white and black female and male high school juniors. Twenty students were observed answering 8 verbal items from one form of the November 1987 SAT. Students were then interviewed about their solution processes.

Subjects

Twenty high school juniors, from four high schools in the Seattle area, participated in this study. The schools were all public high schools, two urban and two suburban in location. All offer above average academic programs, but none serve an elite population. The researcher had difficulty finding black juniors who intended to take the SAT, and was able to find only one black student who had taken the PSAT. The students in the study were unusually well-matched, with comparable GPA's and test coaching experience. The ten males (five black and five white) had an average cumulative GPA of 3.12 and an average cumulative English GPA of 3.24. The ten females (five black and five white) had an average cumulative GPA of 3.12 and an average English GPA of 3.34. Four of the students (two males and two females) had previously received coaching in SAT test-taking. The similarity between males and females in terms of GPA and test preparation permitted exploration of possible differences in solution processing methods related to variables other than gradepoint and level of preparation.

Instrument

Eight items for the study were selected from the Verbal Section of the November 1987 SAT Form Code 7H (see Appendix C for complete text). The questions selected were those items that had the greatest difference in percentage of correct answers between males and females in the "A - A+" grade point average quartile of the Rosser study (The SAT Gender Gap, 1989). This grade point group was selected because it contained the highest achievers and was found to have the largest gender gap on the test.

There were three types of questions that created large differences between males and females with high GPAs: sentence completions, analogies, and antonyms. Each type requires a different sort of processing. Sentence completion questions present a sentence with one or more words missing. They measure a student's ability to recognize words that fit into the context of a sentence both logically and stylistically. Analogy items present one analogous pair of words and ask students to find a matching pair. They measure a student's ability to identify a relationship inherent in a pair of words, and then recognize a similar relationship in a different pair of words. Antonym items present one word and ask students to select its opposite. They measure a student's comprehension of a word and ability to identify its most precise opposite.

TABLE 1

PERCENTAGE OF A+ FEMALES AND A+ MALES ANSWERING ITEM CORRECTLY AS REPORTED IN ROSSER STUDY OF THE NOVEMBER 1987 SAT

Question # (Section-Item)	Question Type	Females % Correct	Males % Correct	Diff
1.(4-23)	Sentence completion (Supreme Court)	64	73	-9
2.(4-24)	Sentence completion (passive listening)	48	57	-9
3.(4-25)	Sentence completion (underdogs)	27	61	-34
4.(1-21)	Analogy (overwhelm:force)	55	63	-8
5.(1-23)	Analogy (dessicate:moisture)	22	29	-7
6.(4-41)	Analogy (dividends:stockholders)	61	75	-14
7.(1-7)	Antonym (prevalent)	44	55	-11
8.(1-8)	Antonym (haphazard)	33	40	-7

These eight items were the ones selected for this study.

Procedure

The selected items were presented to the students with directions and a sample item from the same SAT (see Appendix C.) Students were told to respond to each question as they normally would if they encountered it on the test. They were told that, unlike the actual SAT, there was no time limit for each item and they could take as long as they needed to respond. The interviewer recorded the time for each item, and noted the students' comments as they worked.

After completing each item, they were asked to explain how they arrived at their answer. The interviewer recorded the students' comments on a response form (see Appendix D). The same interviewer was used for all twenty students in the sample.

After completing the questions in a category (sentence completion, analogies, and antonyms), the students were asked to rate the questions for level of difficulty (very easy, easy, not sure, difficult, and very difficult). They were then asked "would you have skipped this question on a real SAT?"

Results

It was hypothesized that the percentages of males and females answering each item correctly would be similar to those reported by the test publisher, Educational Testing Service (ETS) on the November 1987 SAT data tape they had provided for Rosser's research. This hypothesis was partially supported. The profile of differences between males and females in this study was similar to the results from ETS on three of the items (4-25, 4-41, and 1-7), but different on five of the items, as shown on Table 2. The fact that more of the females involved in this study were able to correctly answer these questions suggests that females may be able to respond correctly more frequently outside the pressured test-taking environment.

TABLE 2

PERCENTAGE OF FEMALES AND MALES ANSWERING ITEM CORRECTLY IN THIS STUDY

Question #	Question Type	Females %Correct	Males %Correct	Diff
1.(4-23)	Sentence completion (Supreme Court)	40	40	0
2.(4-24)	Sentence completion (passive listening)	50	40	+10
3.(4-25)	Sentence completion (underdogs)	10	50	-40
4.(1-21)	Analogy (overwhelm:force)	20	20	0
5.(1-23)	Analogy (dessicate:moisture)	30	20	+10
6.(4-41)	Analogy (dividends:stockholders)	40	60	-20
7.(1-7)	Antonym (prevalent)	50	60	-10
8.(1-8)	Antonym (haphazard)	10	10	0

TABLE 3

NUMBER OF STUDENTS ANSWERING CORRECTLY, BY GENDER AND RACE

Item	Subject	Gender	All	White	Black
1 (4-23)	Sentence Completion	F	4	2	2
		M	4	1	3
2 (4-24)	Sentence Completion	F	5	3	2
		M	4	2	2
3 (4-25)	Sentence Completion	F	1	1	0
		M	5	2	3
4 (1-21)	Analogy	F	2	2	0
		M	2	1	1
5 (1-23)	Analogy	F	3	2	1
		M	2	1	1
6 (4-41)	Analogy	F	4	2	2
		M	6	4	2
7 (1-7)	Antonym	F	4	2	2
		M	6	4	2
8 (1-8)	Antonym	F	1	1	0
		M	1	0	1

In general, males responded correctly 30 times out of a possible 80 times, and females responded correctly 24 times. There was also a positive correlation between number of correct responses and GPA, as shown in Table 4.

TABLE 4

CORRELATION BETWEEN NUMBER CORRECT AND GPA

	MALES	FEMALES	BLACKS	WHITES
Total GPA	+.705	+.709	+.603	+.796
English GPA	+.706	+.637	+.551	+.780

It was hypothesized that the time spent on each item by males and females responding correctly would differ. This was found to be true. Females who responded correctly spent more time on 5 of the questions than males who answered correctly, while males who responded correctly spent more time on 3 of the questions than the correctly responding females.

It was also hypothesized that correct responders would use more than the average time allotted within the SAT time limits. This would be an average of 40 seconds per item for section 4 (4-23, 4-24, 4-25, 4-41) and 45 seconds per item for section 1 (1-21, 1-23, 1-7, 1-8). Table 5 shows that on six of the eight items, females responding correctly used more time than the average allowed. Males responding correctly used more time than the average allowed on four of the eight items.

TABLE 5

AVERAGE TIME PER ITEM BY CORRECT RESPONDERS IN SECONDS, BY GENDER

Question #	Question Type	Av.Time In Sec.by Correct Responders		ETS Time Allowed
		Female	Male	Per Item
4-23	Sentence completion (Supreme Court)	69	49.5	40
4-24	Sentence completion (passive listening)	48.8	57.8	40
4-25	Sentence completion (underdogs)	54	44.4	40
1-21	Analogy (overwhelm:force)	44	21	45
1-23	Analogy (dessicate:moisture)	53	95.5	45
4-41	Analogy (dividends:stockholders)	51.3	29	40
1-7	Antonym (prevalent)	28.6	39.1	45
1-8	Antonym (haphazard)	23	7	45

It was hypothesized that the average time used by males and females for answering each item correctly would differ from the average time used by the entire sample of females and males. This was found to be true.

TABLE 6

DIFERENCE IN AVERAGE TIME PER ITEM, TOTAL MALES/FEMALES COMPARED TO MALES/FEMALES ANSWERING ITEM CORRECTLY

Question #	Average Female Time			Average Male Time		
	ALL	CORRECT	DIFF	ALL	CORRECT	DIFF
4-23	51.8	69	+6.2	64.5	49.5	- 15.0
4-24	50.9	48.8	-2.1	45.8	57.8	+ 12.0
4-25	39.7	54	+14.3	48.4	44.4	- 4.0
1-21	38.5	44	+ 5.5	42.8	21	- 21.8
1-23	41.6	53	+11.4	60.7	95.5	+ 34.8
4-41	50.2	51.3	+ 1.1	36.7	29	- 7.7
7-1	38.6	28.3	-10.3	40.7	39.1	- 1.6
1-8	36.6	23	-13.6	30.2	7	- 23.2

The above table illustrates an interesting finding for this small sample: females who answered correctly averaged MORE TIME to respond than the female sample in general, while males who answered correctly averaged LESS TIME than males in general on five of the eight items. A longer processing time for females responding correctly suggests that, given more time or a lack of time pressure on the SAT, females would perform at a higher level.

It was hypothesized from Rosser's study (1989) of questions omitted by females on the November 1987 SAT, that females would be more likely to respond affirmatively to the question "Would you have skipped this item on a real SAT?" This hypothesis was supported, as shown on Table 7. Females indicated they would have omitted 27 (out of 80) responses and males said they would have omitted 15 responses.

TABLE 7

NUMBER OF FEMALES AND MALES INDICATING INTENT TO OMIT ON SAT

Question #	Question Type	Females	Males
4-23	Sentence completion (Supreme Court)	2	2
4-24	Sentence completion (passive listening)	5	2
4-25	Sentence completion (underdogs)	3	0
1-21	Analogy (overwhelm:force)	2	1
1-23	Analogy (dessicate:moisture)	6	6
4-41	Analogy (dividends:stockholders)	2	1
1-7	Antonym (prevalent)	4	2
1-8	Antonym (haphazard)	3	1
	Total Omits:	27	15

When asked this question by the interviewer, one black female responded, "I usually leave them out if I'm not sure." This seems to be a typical response for females. Marcia Linn, et al, (1987) found that females may select the "I don't know" response on a test because of "an unwillingness to take risks." Males, it was theorized, are "expected to take risks. It is better to have guessed and been wrong than to admit uncertainty by saying 'I don't know'." In an earlier study, Linn (1985) concluded that "males are sometimes found to be less realistic than females about estimating their likelihood of success on an academic task and may overrate their confidence in their answers. Thus, they may not realize that their confidence is unjustified and may in fact get the correct answer more often than those who take fewer chances."

Linn's findings were supported in this study as well. Females indicated they would have omitted almost twice as many questions as males if this had been an actual SAT (27 vs. 15). It was also found that females would have answered the same percentage of questions correctly as males, when forced to guess (6 vs. 3).

Females not only reported more intent to omit, but also underestimated their performance in making that decision. Four of the ten females (2 white and 2 black) made negative comments about their ability to perform on the items, while only one male (white) did so. The male, and three of the females, accurately estimated their abilities: they answered either one or none of the items correctly. However, one of the females making the negative comment answered 3 items correctly, which was above average for females as a group.

Related to the above data was an unsolicited response that occurred when the students were asked to indicate whether or not they "would have skipped the item" on the SAT. No directions as to scoring procedures had been given by the interviewer, but when asked to decide whether or not to omit an item, six of the ten males immediately asked whether or not there was a penalty for guessing. Only two of the ten females did so. (The interviewer responded with the standard SAT directive of a 1/4 penalty for incorrect responses.) This difference in behavior between males and females lends support to Harvard Education Professor Carol Gilligan's theory that males are more likely than females to rise to the competitive challenge of the test, viewing it as a game to be won (Rosser, 1989 op cit, page 64.)

An Analysis of Student Performance On Each Question

Student's answer choices, timing, and problem-solving styles were recorded by the interviewer (see Appendix D), creating a response profile for males and females on each item. These profiles are displayed on a chart with each question along with a discussion of the findings. In order to fully characterize the students' performance, a narrative description of the students' response processes is also reported for each of the eight items. A data chart comparing sex, race, number of correct responses, GPA's, and test-wiseness for the entire sample may be found in Appendix A.

Question 1 (4-23)

Complete the following sentence:

The ruling by the Supreme Court that the appropriation by a state government of Cherokee lands was unconstitutional did not --- such seizures because President Jackson failed to --- it.

(A) legitimize..ratify (B) change..explicate (C) justify..challenge
(D) end..commute (E) halt..implement CORRECT RESPONSE: E

FEMALE RESPONSE PROFILE

RESPONSE PROFILE Question #4-23	5 WHITE FEMALES				5 BLACK FEMALES					
A. Response chosen * X=correct choice	A	X	X	C	A	X	D	D	B	X
B. Time elapsed in seconds	51	70	62	23	35	57	42	39	52	87
C. Difficulty rating One=Easy;Five=Difficult	4	2	4	2	4	4	4	3	4	4
D. Would skip this item on actual SAT	0	0	1	0	0	0	0	0	1	0
E. Percent of vocabulary known	90	90	90	80	80	80	50	60	80	100
F. Attempted to decipher unknown words	0	1	0	0	0	1	1	0	0	*
G. Eliminated choices during response selection	1	1	1	1	1	1	1	0	0	1
H. Used pre-planned system for response selection	1	0	1	0	0	0	0	0	0	1
I. Able to explain how response was selected	1	1	1	1	1	1	1	0	1	1
J. Had previous familiarity of subject matter	0	1	1	0	1	0	1	0	0	0
K. Response decision based on first word of choice pair	0	0	0	1	0	0	1	0	1	0
L. First determined meaning for blanks, then sought match	0	0	1	0	1	1	1	0	0	1

*Letter of incorrect response chosen or X for correct choice

MALE RESPONSE PROFILE

RESPONSE PROFILE Question #4-23	5 WHITE MALES					5 BLACK MALES				
A. Response chosen X=correct choice	C	A	A	X	C	X	X	C	C	X
B. Time elapsed in seconds	35	62	97	49	79	29	40	150	24	80
C. Difficulty rating One=Easy;Five=Difficult	4	5	4	4	4	3	3	4	4	4
D. Would skip this item on actual SAT	0	1	0	0	1	0	0	0	0	0
E. Percent of vocabulary known	80	90	80	90	80	90	90	90	80	70
F. Attempted to decipher unknown words	0	0	0	0	1	0	0	0	0	1
G. Eliminated choices during response selection	0	1	1	1	1	1	1	1	0	1
H. Used pre-planned system for response selection	0	0	1	1	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	1	1	1	1	1	1	1
J. Had previous familiarity of subject matter	0	0	1	1	1	1	1	0	0	1
K. Response decision based on first word of choice pair	1	0	0	0	1	0	0	1	1	0
L. First determined meaning for blanks, then sought match	0	0	0	1	0	1	1	0	0	1

Discussion

Forty percent of the males (1 white, 3 black) and forty percent of the females (2 white, 2 black) answered this item correctly. Knowledge of vocabulary seemed to be an important factor in success on this item. Of the eight students answering correctly, one knew 100% of the words and 6 knew 90%. The others knew 80% and 70% respectively, but also made attempts to decipher unknown words.

One student commented, "I try to break a sentence down into my own words - I couldn't do it with this one because I didn't know appropriations." (However, she made a good guess at the word and answered the item correctly.)

Students made up their own summary sentences:

"If he failed to ratify it it wouldn't be legitimate"

and thought of words that could go in the blanks:

"I knew it had to be - did not STOP the procedures - halt and end would work, but commute was wrong"..."ratify sounded right for the last part but legitimate wasn't right - I saw halt and knew that was right."

All successful respondents eliminated responses during the process. Six of the males and four of the females reported prior knowledge of the topic. All of the correct male respondents reported prior knowledge; only two of the four correct female reported this. Four females chose distractor (C) justify..challenge, and the remaining two chose distractor (A) legitimize..ratify. Males showed no preference in selecting distractors.

Question 2 (4-24)

Complete the following sentence.

When people habitually read or study with music playing in the background, they unfortunately cultivate "passive listening," which is the very --- of listening, a process that is essentially active.

(A) presage (B) antithesis (C) circumstance (D) pinnacle
(E) illumination CORRECT RESPONSE: B

FEMALE RESPONSE PROFILE

QUESTION 4-24	5 WHITE FEMALES					5 BLACK FEMALES				
A. Response chosen X=correct choice	X	D	A	X	X	D	X	X	A	A
B. Time elapsed in seconds	51	45	52	25	42	57	28	98	52	59
C. Difficulty rating One=Easy;Five=Difficult	4	5	5	4	5	5	3	4	4	2
D. Would skip this item on actual SAT	0	1	0	1	1	1	0	0	1	0
E. Percent of vocabulary known	60	40	60	40	40	40	40	40	40	60
F. Attempted to decipher unknown words	1	1	1	0	1	0	1	0	0	0
G. Eliminated choices during response selection	1	1	1	0	0	1	1	1	1	1
H. Used pre-planned system for response selection	1	0	1	0	0	0	0	0	0	1
I. Able to explain how response was selected	1	1	1	0	1	1	1	0	1	1
J. Had previous familiarity of subject matter	0	0	0	0	1	0	0	0	0	0
K. Response decision based on first word of choice pair	*	*	*	*	*	*	*	*	*	*
L. First determined meaning for blanks, then sought match	1	0	1	0	1	0	1	0	0	1

MALE RESPONSE PROFILE

QUESTION 4-24	5 WHITE MALES					5 BLACK MALES				
A. Response chosen X=correct choice	D	E	X	X	D	X	A	X	E	D
B. Time elapsed in seconds	30	34	57	26	32	48	42	100	26	63
C. Difficulty rating One=Easy;Five=Difficult	3	3	4	3	2	4	5	4	4	5
D. Would skip this item on actual SAT	0	0	0	0	0	0	1	1	0	0
E. Percent of vocabulary known	60	20	60	60	60	60	60	80	40	60
F. Attempted to decipher unknown words	0	0	1	1	0	1	1	1	0	0
G. Eliminated choices during response selection	1	0	1	1	1	1	1	1	1	1
H. Used pre-planned system for response selection	0	0	1	1	0	0	1	0	0	0
I. Able to explain how response was selected	1	0	1	1	1	1	1	1	0	1
J. Had previous familiarity of subject matter	0	0	0	0	0	0	0	0	0	0
K. Response decision based on first word of choice pair	*	*	*	*	*	*	*	*	*	*
L. First determined meaning for blanks, then sought match	0	0	1	1	1	1	1	1	0	0

Fifty percent of the females (3 white, 2 black) and forty percent (2 white, 2 black) of the males answered this item correctly. Only 1 of the 10 males, and 1 of the 10 females were familiar with the word "antithesis." But the important factor in success on this item was the attempt to decipher unknown words. Of the nine students answering correctly, one knew 80% of the words, 4 knew 60%, and the remaining four knew 40%. However, 7 of the 9 correct responders attempted to decipher the unknown words, one of which was "antithesis," the correct answer. The process of reasoning was excellent in many instances. A female student said:

> "It's not good to listen to music when you study, so it had to be something that takes away...anti means against..that's negative...it means a loss"

Another said:

> "It had to be the opposite of listening and anti means opposite."

Incorrect reponders either missed the sense of the passage:

> "It has to mean the same thing as passive."

or selected a word they did not know in hopes it would be correct:

> "It means the very opposite of listening .. I figured that was what presage meant".

All successful male respondents eliminated responses during the process, while only three of the five successful females did so. None of the males and only one of the females reported prior knowledge of the concept of passive listening (she answered correctly). Three

females chose (A) presage, as their answer - although none of them knew what it meant - and three males chose (D) pinnacle, apparently in an attempt to replicate a familiar syntactic pattern "the very pinnacle."

Question 3 (4-25)

Complete the following sentence.

Although the undefeated visitors --- triumphed over their underdog opponents, the game was hardly the --- sportswriter had predicted.

(A) fortunately..upset (B) unexpectedly..classic

(C) finally..rout (D) easily..stalemate (E) utterly..mismatch

CORRECT RESPONSE: C

FEMALE RESPONSE PROFILE

QUESTION 4-25	5 WHITE FEMALES					5 BLACK FEMALES				
A. Response chosen X=correct choice	X	D	D	B	E	D	D	D	D	D
B. Time elapsed in seconds	54	90	13	19	42	35	41	35	50	18
C. Difficulty rating One=Easy;Five=Difficult	4	5	4	4	4	2	3	3	4	1
D. Would skip this item on actual SAT	1	1	0	0	0	0	0	0	1	0
E. Percent of vocabulary known	60	80	90	80	70	80	80	80	90	90
F. Attempted to decipher unknown words	0	0	0	0	0	0	0	0	0	0
G. Eliminated choices during response selection	1	1	1	1	1	1	1	0	0	1
H. Used pre-planned system for response selection	1	0	0	0	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	0	1	1	1	1	1	1
J. Had previous familiarity of subject matter	1	0	0	1	0	1	0	0	0	0
K. Response decision based on first word of choice pair	0	1	1	1	0	1	1	1	1	1
L. First determined meaning for blanks, then sought match	1	0	0	0	1	0	0	0	0	0

MALE RESPONSE PROFILE

QUESTION 4-25	5 WHITE MALES					5 BLACK MALES				
A. Response chosen X=correct choice	D	A	X	X	E	X	X	D	B	X
B. Time elapsed in seconds	40	47	34	52	59	29	56	90	26	51
C. Difficulty rating One=Easy;Five=Difficult	2	2	2	2	1	4	3	3	3	2
D. Would skip this item on actual SAT	0	0	0	0	0	0	0	0	0	0
E. Percent of vocabulary known	90	90	100	100	100	100	100	90	90	100
F. Attempted to decipher unknown words	0	0	*	*	*	*	*	0	0	*
G. Eliminated choices during response selection	0	1	1	1	0	1	1	0	1	0
H. Used pre-planned system for response selection	0	0	1	1	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	1	1	1	1	1	1	1
J. Had previous familiarity of subject matter	0	1	1	1	0	1	1	0	1	1
K. Response decision based on first word of choice pair	1	1	0	0	1	0	0	0	1	0
L. First determined meaning for blanks, then sought match	0	0	1	1	0	1	1	0	0	1

Discussion

Ten percent of the females (one white) and fifty percent of the males (2 white, 3 black) answered this item correctly. This question showed the largest gender difference on the November 1987 SAT and the females performed just as badly in this small study as females did nationwide (25% more males than females answered correctly in a national sample of 100,000.) Students were asked to repond to the question "are you familiar with the sports world?" All who responded correctly, including the one female, replied "yes." The three males who answered "no" to this question also missed the item. Familiarity with the sports world is undoubtedly linked to knowledge of sports vocabulary and both were critical success factors on this item. Most females did not know the meaning of the correct answer "rout", nor did they know the word "underdog."

Instead, seven females chose distractor (D) "easily," basing their response on the word "undefeated" in the stem sentence. They selected the response with the first word "easily" since, in their words:

> "...if they were undefeated they would easily win."

Most females were not sure of the meaning of the word "stalemate," but felt secure in their answer because of the first word. As one female put it:

> "If they were undefeated they wouldn't fortunately or unexpectedly win - those can't be right...finally seems like they've been trying and trying - doesn't make sense - easily seemed best... I'm not sure what stalemate means, but it's the best choice because of the first word."

This logic is very sound and the item is clearly biased in favor of those who know the language of sports. The one female answering correctly did know the meaning of "rout" and "underdog." Predictably, she also indicated that she was "familiar with the sports world." None of the 20 students made any attempts to decipher unknown words.

Question 4 (1-21)

Select the pair of words whose relationship is most similar to the relationship of the stem pair.

OVERWHELM : FORCE : :

(A) conquer : victory

(B) relieve : pressure

(C) entrap : deceit

(D) surrender : strength

(E) bully : submission

CORRECT RESPONSE: C

FEMALE RESPONSE PROFILE

QUESTION 1-21	5 WHITE FEMALES					5 BLACK FEMALES				
A. Response chosen X=correct choice	X	D	X	B	D	B	B	B	B	A
B. Time elapsed in seconds	68	29	24	21	35	82	44	45	15	22
C. Difficulty rating One=Easy;Five=Difficult	3	3	4	5	4	4	4	4	4	1
D. Would skip this item on actual SAT	0	1	0	0	1	0	0	0	0	0
E. Percent of vocabulary known	100	67	100	75	91	100	100	91	100	100
F. Attempted to decipher unknown words	*	1	*	0	0	*	*	0	*	*
G. Eliminated choices during response selection	1	1	1	0	1	1	1	0	0	1
H. Used pre-planned system for response selection	1	0	1	0	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	0	1	1	1	0	1	1
J. Could explain concept of analogy	1	1	1	1	1	0	1	0	0	0
K. Response decision based on horizontal comparisons	1	1	1	0	1	0	1	0	0	1
L. Response decision based on vertical comparisons	0	0	0	1	0	1	0	0	1	0
M. First made pairs into sentences, then sought match	1	1	1	0	1	0	1	0	0	1

MALE RESPONSE PROFILE

QUESTION 1-21	5 WHITE MALES					5 BLACK MALES				
A. Response chosen X=correct choice	A	E	D	X	E	A	X	A	B	A
B. Time elapsed in seconds	16	77	43	26	48	47	16	48	37	70
C. Difficulty rating One=Easy;Five=Difficult	4	4	2	4	2	4	1	3	3	4
D. Would skip this item on actual SAT	0	1	0	0	0	0	0	0	0	0
E. Percent of vocabulary known	83	91	100	100	100	100	100	100	91	83
F. Attempted to decipher unknown words	0	0	*	*	*	0	*	*	0	0
G. Eliminated choices during response selection	1	1	1	1	1	0	0	1	0	0
H. Used pre-planned system for response selection	0	0	1	1	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	1	1	1	1	1	1	1
J. Could explain concept of analogy	1	1	1	1	1	0	1	1	0	0
K. Response decision based on horizontal comparisons	1	1	1	1	1	0	1	1	0	0
L. Response decision based on vertical comparisons	0	0	0	0	0	1	0	0	1	1
M. First made pairs into sentences, then sought match	1	0	1	1	1	0	1	1	0	0

Discussion

Twenty percent of the females (2 white) and twenty percent of the males (1 white, 1 black) answered this item correctly. In all four instances, correctly responding students indicated a 100% knowledge of vocabulary and based their response on horizontal comparisons; that is, they established a relationship between the words in each pair, rather than comparing individual words vertically on the right and left sides. All successful respondents used a process wherein they first made the stem pair into a sentence which clarified the relationship between the two words, and then systematically tried each of the choice pairs in the sentence frame they had constructed:

"He was overwhelmed by force; he was entrapped by deceit."

"When you overwhelm someone you use a lot of force; when you entrap someone you use a lot of deceit."

Error responses by other students were related to poor utilization of this technique:

"When someone is overwhelmed it leads to force; when you conquer something it leads to victory"

"When you overwhelm somebody you force them; when you bully someone, they submit"

"They were overwhelmed because of the force; they surrendered because of the strength."

All correctly responding students were able to explain what an analogy is. All 5 white females and all 5 white males were able to make this explanation, but only one black female and two black males could do so.

These patterns remained the same for all three analogy items. Black students, both male and female, tended to make vertical comparisons, comparing words vertically between the stem and the options, on either side of the pairs. One white and four black females chose distractor (B) "relieve : pressure" because, in their words:

"Overwhelm and relieve have similar meanings; so do force and pressure".

One white and three black males chose distractor (A) conquer..victory:

"You overwhelm someone like you conquer them".

This pattern concurs with a study of factors affecting differential item functioning for Black examinees on the SAT analogy items by Schmitt and Bleistein (ETS, 1987) which found that: " In general, a vertical type of strategy seems to be more consistently used by Black students. Of all of the factors investigated, vertical relationships provided the clearest and most consistent explanation for differential item functioning on analogy items."

Question 5 (1-23)

Select the pair of words whose relationship is most similar to the relationship of the stem pair.

DESSICATE : MOISTURE : :

(A) displace : volume

(B) engender : originality

(C) implicate : logic

(D) dispirit : morale

(E) innundate : fluid

CORRECT RESPONSE: D

FEMALE RESPONSE PROFILE

QUESTION 1-23	5 WHITE FEMALES					5 BLACK FEMALES				
A. Response chosen X=correct choice	X	X	E	C	A	E	X	E	E	E
B. Time elapsed in seconds	45	55	23	12	30	28	59	28	42	94
C. Difficulty rating One=Easy;Five=Difficult	4	3	5	5	5	5	5	4	4	4
D. Would skip this item on actual SAT	0	0	1	1	1	1	1	0	0	1
E. Percent of vocabulary known	91	75	83	50	75	67	67	58	58	75
F. Attempted to decipher unknown words	1	1	1	0	0	0	1	0	0	1
G. Eliminated choices during response selection	1	1	1	0	1	0	0	0	0	0
H. Used pre-planned system for response selection	1	0	1	0	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	0	1	0	1	1	1	1
J. Could explain concept of analogy	1	1	1	1	1	0	1	0	0	0
K. Response decision based on horizontal comparisons	1	1	1	0	1	0	1	0	0	0
L. Response decision based on vertical comparisons	0	0	0	0	0	1	0	1	1	1
M. First made pairs into sentences, then sought match	1	1	0	0	0	0	0	0	0	0

MALE RESPONSE PROFILE

QUESTION 1-23	5 WHITE MALES					5 BLACK MALES				
A. Response chosen X=correct choice	A	C	X	B	A	A	na	X	E	D
B. Time elapsed in seconds	27	61	105	85	59	36	51	86	16	81
C. Difficulty rating One=Easy;Five=Difficult	5	5	5	5	4	4	5	4	3	5
D. Would skip this item on actual SAT	1	1	1	0	1	1	1	0	0	0
E. Percent of vocabulary known	33	67	75	83	75	83	75	75	83	58
F. Attempted to decipher unknown words	0	0	1	1	1	0	0	1	0	0
G. Eliminated choices during response selection	0	0	1	1	0	0	0	1	0	1
H. Used pre-planned system for response selection	0	0	1	1	0	0	0	0	0	0
I. Able to explain how response was selected	0	0	1	1	1	0	0	1	1	1
J. Could explain concept of analogy	1	1	1	1	1	0	1	1	0	0
K. Response decision based on horizontal comparisons	0	0	1	1	1	0	0	1	0	0
L. Response decision based on vertical comparisons	0	0	0	0	0	0	0	0	1	1
M. First made pairs into sentences, then sought match	0	0	1	0	0	0	0	1	0	0

Discussion

Thirty percent of the females (2 white, 1 black) and twenty percent of the males (1 white, 1 black) answered this item correctly. These students had an average vocabulary knowledge level of 76%, and all made attempts to decipher unknown words. All based their response on horizontal comparisons. It is interesting to note that none of the twenty students were familiar with the word DESSICATE, yet five of them still were able to answer the item correctly:

"I think dessicate means to lessen - take away from"

"Dessicate - like desegregate - to take apart, to break down"

"Dessicate - to be ruined, eroded"

"It looks like a negative word - it might mean to dry out"

All but one of the successful respondents made the stem pair into a sentence which clarified the relationship between the two words, and then systematically tried each of the choice pairs in the sentence frame they had constructed.

"I thought dessicate means to take away from your moisture; dispirit means to take away from your morale"

Five females (1 white and 4 black) chose the distractor choice (E) because of horizontal comparisons between "moisture" and "fluid":

"Moisture is almost the same as fluid"

"Fluid has something to do with moisture"

But two students, rejecting (E) in favor of the correct response, commented:

"They're trying to trick you with something to do with water" (white male")

"That's too obvious - they don't make them that easy" (white female)

Question 6 (4-41)

Select the pair of words whose relationship is most similar to the relationship of the stem pair.

DIVIDENDS : STOCKHOLDERS ::

(A) investments : corporations

(B) purchases : customers

(C) royalties : authors

(D) taxes : workers

(E) mortgages : homeowners

CORRECT RESPONSE: C

FEMALE RESPONSE PROFILE

QUESTION 4-41	5 WHITE FEMALES					5 BLACK FEMALES				
A. Response chosen X=correct choice	E	A	X	E	X	B	X	E	B	X
B. Time elapsed in seconds	53	82	51	40	28	28	16	15	79	110
C. Difficulty rating One=Easy;Five=Difficult	3	3	5	2	2	4	3	3	4	2
D. Would skip this item on actual SAT	1	0	0	0	0	1	0	0	0	0
E. Percent of vocabulary known	100	100	100	91	91	91	91	83	100	100
F. Attempted to decipher unknown words	*	*	*	0	0	0	1	0	*	*
G. Eliminated choices during response selection	0	1	1	1	1	1	0	0	1	1
H. Used pre-planned system for response selection	0	0	1	0	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	1	1	1	1	1	1	1
J. Could explain concept of analogy	1	1	1	1	1	0	1	0	0	0
K. Response decision based on horizontal comparisons	1	1	1	1	1	0	1	0	0	1
L. Response decision based on vertical comparisons	0	0	0	0	0	1	0	1	1	0
M. First made pairs into sentences, then sought match	1	1	1	1	1	0	1	0	0	1

MALE RESPONSE PROFILE

QUESTION 4-41	5 WHITE MALES					5 BLACK MALES				
A. Response chosen X=correct choice	X	B	X	X	X	E	X	X	A	A
B. Time elapsed in seconds	17	69	59	31	19	68	14	34	10	46
C. Difficulty rating One=Easy;Five=Difficult	2	2	4	2	1	4	2	1	3	4
D. Would skip this item on actual SAT	0	0	0	0	0	1	0	0	0	0
E. Percent of vocabulary known	100	91	91	100	100	83	100	100	100	91
F. Attempted to decipher unknown words	*	0	1	*	*	0	*	*	*	0
G. Eliminated choices during response selection	1	1	1	1	1	1	0	0	0	0
H. Used pre-planned system for response selection	0	0	1	1	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	1	1	1	1	1	1	1
J. Could explain concept of analogy	1	1	1	1	1	0	1	1	0	0
K. Response decision based on horizontal comparisons	1	0	1	1	1	1	1	1	0	0
L. Response decision based on vertical comparisons	0	1	0	0	0	0	0	0	1	1
M. First made pairs into sentences, then sought match	1	0	1	1	1	1	1	1	0	0

Discussion

Forty percent of the females (2 white and 2 black) and sixty percent of the males (4 white, 2 black) answered this item correctly. In all instances, correctly responding students indicated either a 100% or 91% knowledge of vocabulary. All correct students based their response on horizontal comparisons. All successful respondents first made the stem pair into a sentence and then tried each of the choice pairs in the sentence frame they had constructed:

> "If you hold stock you get money; if you write a book you get money"
>
> "Stockholders are getting something back out of what they've done; authors are getting something back out of what they've done."

Others tried to make sentences, but were unclear as to the relationship between the stem words:

> "I tried to decide if dividends were something they got or something they gave"

One student constructed the sentence "Dividends are bought by stockholders" and then checked each choice pair. When nothing fit, she made up a different sentence: "Stockholders receive dividends, which are beneficial." She was then able to match it with: "Authors receive royalties, which are beneficial," and responded correctly.

Average overall vocabulary knowledge for this item was high: 94.7% for females and 95.5% for males. However, four females (3 black, 1 white) and two males (1 black, 1 white) did not know the meaning of the word "dividends," and two females (black) and two males (1 black, 1 white) did not know the meaning of the word "royalties," which was part of the correct reponse.

Again, black students tended to make vertical comparisons:

"Dividends are like purchases and stockholders are like customers."

"I think dividends are sort of like investments, and corporations have stockholders"

Question 7 (1-7)

Select the word which is most nearly opposite in meaning to the stem word.

PREVALENT

(A) limited (B) confused (C) inexact (D) important (E) healthy

CORRECT RESPONSE: A

FEMALE RESPONSE PROFILE

QUESTION 1-7	5 WHITE FEMALES					5 BLACK FEMALES				
A. Response chosen * X=correct choice	X	X	C	C	X	B	X	C	D	X
B. Time elapsed in seconds	19	58	59	22	6	75	20	48	39	40
C. Difficulty rating One=Easy;Five=Difficult	3	3	5	4	2	5	2	4	5	2
D. Would skip this item on actual SAT	1	0	1	0	0	0	0	1	1	0
E. Percent of vocabulary known	100	100	82	66	100	66	100	82	82	82
F. Attempted to decipher unknown words	*	*	1	0	*	0	*	0	0	1
G. Eliminated choices during response selection	1	1	1	1	0	0	0	0	0	1
H. Used pre-planned system for response selection	1	0	0	0	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	1	1	0	1	0	0	1
J. Could give correct meaning of stem word	1	1	0	0	1	0	1	0	0	1
K. Found opposites for choices then sought match to stem	0	0	0	0	0	0	0	0	0	0
L Found opposite for stem sought match from choices	1	0	1	0	1	0	1	0	0	1

MALE RESPONSE PROFILE

QUESTION 1-7	5 WHITE MALES					5 BLACK MALES				
A. Response chosen * X=correct choice	B	X	X	X	C	X	X	B	X	D
B. Time elapsed in seconds	24	94	66	10	54	11	14	76	20	38
C. Difficulty rating One=Easy;Five=Difficult	4	4	4	4	3	2	2	3	4	5
D. Would skip this item on actual SAT	0	1	0	0	0	0	0	0	0	1
E. Percent of vocabulary known	82	100	100	100	82	100	100	82	100	66
F. Attempted to decipher unknown words	1	*	*	*	1	*	*	1	*	0
G. Eliminated choices during response selection	1	1	1	1	1	0	0	1	0	0
H. Used pre-planned system for response selection	0	0	1	1	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	1	1	1	1	1	1	0
J. Could give correct meaning of stem word	0	1	1	1	0	1	1	0	1	0
K. Found opposites for choices then sought match to stem	1	0	1	0	0	0	0	0	0	0
L. Found opposite for stem sought match from choices	0	1	0	1	1	1	1	1	1	0

Discussion

Fifty percent of the females (3 white, 2 black) and sixty percent of the males (3 white, 3 black) answered this item correctly. The nine students who answered incorrectly were not familiar with the word "prevalent." Knowledge of vocabulary was an important factor in success on this item, combined with attempts to decipher "prevalent" when the student was only partially sure of the meaning. Of the eleven students answering correctly, three used their memory of a similar word, "prevailing," to assist them in determining the meaning of "prevalent."

Most of the correct reponders determined the opposite for "prevalent" in their own minds and then searched for a match among the choices. Two of the correct responders, however (both white males), thought of an opposite for each of the choices and then, having thought of "unlimited" as the opposite for "limited," decided it was closest to "prevalent" and selected it as their answer.

The most common distractor selected by females was (C) "inexact". The most common distractor selected by males was (B) "confused."

Question 8 (1-8)

Select the word which is most nearly opposite in meaning to the stem word.

HAPHAZARD

(A) resistant (B) original (C) secure (D) considered

(E) confidential

CORRECT RESPONSE: C

FEMALE RESPONSE PROFILE

QUESTION 1-8	5 WHITE FEMALES					5 BLACK FEMALES				
A. Response chosen X=correct choice	C	A	X	D	C	A	C	C	C	C
B. Time elapsed in seconds	62	39	23	26	6	64	6	8	65	67
C. Difficulty rating One=Easy;Five=Difficult	3	3	3	4	2	4	2	4	4	4
D. Would skip this item on actual SAT	0	0	0	0	0	0	0	1	1	1
E. Percent of vocabulary known	100	82	100	82	82	82	100	82	82	82
F. Attempted to decipher unknown words	*	1	*	1	0	1	*	0	1	1
G. Eliminated choices during response selection	1	1	0	1	0	1	0	0	0	0
H. Used pre-planned system for response selection	1	0	0	0	0	0	0	0	0	0
I. Able to explain how response was selected	1	1	1	0	1	1	1	0	1	1
J. Could give correct meaning of stem word	1	0	1	0	0	0	1	0	0	0
K. Found opposites for choices then sought match to stem	1	0	0	0	0	0	0	0	0	0
L. Found opposite for stem sought match from choices	0	1	1	0	1	0	0	0	1	1

MALE RESPONSE PROFILE

QUESTION 1-8	5 WHITE MALES					5 BLACK MALES				
A. Response chosen X=correct choice	C	C	C	C	C	C	X	C	C	C
B. Time elapsed in seconds	6	9	29	31	26	16	7	85	30	63
C. Difficulty rating One=Easy;Five=Difficult	2	1	2	4	2	3	2	5	4	4
D. Would skip this item on actual SAT	0	0	0	0	0	0	0	1	0	0
E. Percent of vocabulary known	82	82	82	100	82	100	100	82	82	82
F. Attempted to decipher unknown words	0	0	0	*	0	*	*	1	1	0
G. Eliminated choices during response selection	0	0	1	1	1	0	0	0	0	0
H. Used pre-planned system for response selection	0	0	1	1	0	0	0	0	0	0
I. Able to explain how response was selected	1	0	1	1	1	1	1	0	1	0
J. Could give correct meaning of stem word	0	0	0	1	0	1	1	0	0	0
K. Found opposites for choices then sought match to stem	0	0	1	0	0	0	0	0	0	0
L. Found opposite for stem sought match from choices	0	1	0	1	1	1	1	0	1	1

Discussion

Only one white female and one black male successfully answered this item. Sixty percent of the females, and ninety percent of the males chose distractor (C) "secure" instead. This apparently occurred because students inserted an element of danger into their definition of "haphazard." They defined it as meaning "risky" or "unsafe," probably due to the sub-word "hazard," which all student recognized. Having decided the word meant something dangerous, they then chose "secure" as the best opposite.

One correct responder (white female) again thought of an opposite for each of the choices and then, having given "unconsidered" as the opposite for "considered," decided it was closest to "haphazard" and selected it as her response. It is interesting to note that this student had received coaching on how to take the SAT, and indicated that she had been taught this method.

SUMMARY OF MAJOR FINDINGS

Cognitive Differences: There were no major differences between the sexes in problem-solving styles used to obtain both correct and incorrect responses. However, there was a difference between white and black students in solving analogy questions that supports findings by ETS researchers Schmitt and Bleistein. Black students were more likely to look for vertical comparisons between words on both sides of the analogy pairs, rather than matching the horizontal relationship between the word pairs, which was the skill being tested.

Correct Responders: Equal numbers of white females, white males, and black males responded correctly. Black females' correct responses were 60% of the other three groups. Black females performed less well than the other groups, even though they had received a comparable education. Students who answered correctly shared several characteristics:

They knew more of the vocabulary.

They were more likely to try to decipher unknown words.

They were more systematic about making decisions.

They had a more strategic approach to finding the correct answer, and changed strategies depending on the type of question (sentence completion, analogy, or antonym)

There were no significant differences between the sexes in exhibiting these characteristics.

Timing: Females who answered correctly averaged more time to respond than males. They also averaged more time per item than would be allowed by ETS under regular testing conditions.

Risk-Taking: Females said they would omit items almost twice as often as males (27 as compared to 15). Yet, when they did answer them, they got approximately the same percentage correct as males (20%). Females were also less likely to ask about scoring rules when considering whether to omit.

Prior Knowledge: Previous familiarity with the subject matter of a question was helpful in enabling students to respond correctly. When topics are of greater interest to males (i.e. sports, stock market), they perform better on those items than females.

Possible Effects of Coaching: Students who had received coaching had more sophisticated test-taking strategies. The two females and two males who had been coached used systematic procedures for each item type: determining meanings for blanks on sentence completions, making word pairs into sentences on analogies, finding opposites for each response choice on antonyms, and quickly eliminating one or more answers to narrow the field. And they got 26% more questions correct than the sample in general.

SUGGESTIONS FOR FURTHER RESEARCH

The main suggestion for further research is that this study be replicated with a larger number of students, to establish whether the findings here also apply to the national test-taking pool.

This small sample uncovered several interesting cognitive patterns which merit further study with a larger number of male and female, and/or black and white students to determine whether differences in mental processes are creating gender and racial differences on the test. Little research has been done outside of this study on the components that make up the reasoning tasks required by these verbal items from the SAT.

It seems clear from this study that black students process analogies differently than whites, with resulting errors in their responses. Additional study of this phenomenon might include testing experimental groups of black students who have received coaching in the process of analogical reasoning to determine if the outcomes can be impacted through intervention.

Research that focuses on specific components of the solution process may yield further information on differences between males and females or blacks and whites. Cognitive styles expert Dr. Robert Sternberg, Professor of Psychology at Yale University, divides verbal tasks into the following processes:

comprehending the terms of the stem and the task required

inferring the relationship which should occur (missing word, analogous relationship, opposite relationship)

applying the results of the inference to generate an ideal answer

justification/elimination process which selects among alternative answers, none of which precisely matches the ideal answer

selecting the response judged best of the options presented

Additional studies might identify gender or racial performance differences on these verbal processes.

The entire area of "test-wiseness" merits further study. Both male and female students who had received coaching prior to participating in the study demonstrated test-taking strategies that enabled them to perform at a higher level than students without this kind of instruction. It would be interesting to set up matched pairs of females and males, provide coaching to one member of each pair, and then document their test-taking behaviors.

The time factor should be looked at in greater detail. This study seems to indicate that, given unlimited time on the SAT, it is possible that females would select a greater number of correct responses than males. Current practices in the field of higher order thinking skills focus on decreasing "impulsivity" in problem-solving situations, yet the SAT rewards the student who can complete the greatest number of items within tight time constraints.

The tendency for females to omit answers is another area that merits further study. This study suggests that females are less likely than males to risk the possibility of an incorrect answer. This may be related to females' desire to comply with the stated

guidelines regarding penalties for incorrect answers. It would be interesting to give the SAT to a group of males and females after telling them that the penalty for guessing had been removed (and then scoring it without penalties.) Additional research could be done in the area of female and male response to competitive game situations and the extent to which females and males perceive testing situations as "gamelike." Specific behaviors on the test could be compared to students' behaviors in competitive gaming situations.

And finally, the entire issue of item bias needs to be looked at in depth. The correlation between prior knowledge of a subject area and correct response on an item from that subject area is substantial. When members of one group (female) must take a test which includes items that reflect special interest areas (sports, finance) of a different group (male), they are obviously at a disadvantage in competing with that other group. Attempting to balance the test with math items on oatmeal cookies is more of an insult than a solution.

AN INVESTIGATION OF THE PERFORMANCE DIFFERENCES OF

WHITE AND BLACK FEMALE AND MALE HIGH SCHOOL STUDENTS

ON FIVE MATH QUESTIONS FROM THE NOVEMBER 1987 SCHOLASTIC APTITUDE TEST

The purpose of this exploratory study was to investigate the ways that various students process and solve SAT math questions. The study investigated the solution methods employed by white and black female and male high school juniors. Forty students were observed solving 5 SAT math questions from one form of the November of the 1987 SAT math test. Students were then interviewed about their performance, view of the test problems, and attitudes.

Methods

Subjects

Forty high school juniors participated in this study. The students were solicited from two New York City public high schools, a New York City Catholic school, and a test preparation course. Most of the students enrolled in the test preparation course attended private New York City high schools.

Students from the three schools were enlisted by the interviewers. First, the interviewers, using school records, identified students that had the appropriate qualifications (B+ or better in Algebra and Geometry and a PSAT math score of at least 43). These students were then asked if they would like to participate. Students from the test prep program were enlisted through a questionnaire mailed to all students enrolled in a Princeton Review coaching course.

From this pool of students, 20 females (10 black and 10

white) and 20 males (10 black and 10 white) were chosen to take part in the study. These students all met the criteria stated above (see Table 1). Students who completed the interviews received either $10.00 or a magazine subscription as compensation.

Table 1

Average Preliminary Scholastic Aptitude Test (PSAT) scores and grade point averages (GPA) for participating students, by race and gender.

		MPSAT	VPSAT	GPA
Females	White	55.6	51.7	89.1
	Black	47.2	47.3	86.9
Males	White	51.4	46.6	82.9
	Black	51.7	41.1	82.4

N = 40 (20 females and 20 males)

The selection of students with average or above average PSAT math scores and similar mathematical knowledge allowed for the exploration of possible differences in girls' and boys' solution methods, rather than in their mathematical knowledge. According to Educational Testing Service (ETS), the average PSAT math scores for 1987 were 43.1 for all females and 47.2 for all males.

The 20 females participating in the study had an average PSAT math score of 51.4 and the 20 males had an average PSAT math score of 51.5. However, as seen in Table 1, the average PSAT math averages for the white and black females were very different. The white females had an average PSAT math score

of 55.6 whereas the black females had a PSAT math average of 47.2. The average PSAT math scores for the white and black males were similar, 51.4 and 51.7.

Instrument

Five questions were selected from the November 1987 SAT math test form code 7H (see Appendix E). The questions were selected from those items that had at least a 12% difference between females and males, for both the "A+/A" and "A-/B+" GPA groups1 (see Table 2). Further the selected items had differences in the distractor selections for females and males (see Table 3).

Table 2

Percentage of students answering correctly, by gender, grade point average (GPA), and race, for each of the selected SAT items, as reported by ETS.

				GPA[1]			
Item No.	Subject	Gender	All	A+/A	A-/B+	White	Black
1	Algebra	F	42	66	47	44	25
		M	52	79	61	53	32
2	Arithmetic	F	24	45	28	25	13
		M	33	62	42	35	17
3	Geometry	F	58	77	65	61	38
		M	74	90	82	77	51
4	Geometry	F	47	72	56	50	25
		M	60	85	72	62	34
5	Arithmetic	F	35	54	39	37	20
		M	46	70	53	49	24

N = 99997 (92984 reporting GPA) F = 54606 (51242 with GPA)
M = 45391 (41742 with GPA)

[1] The GPA groups were based on ETS's data of students' self-reported GPA.

Table 3

Percentage of females and males responding to each alternative including omits.

Item No.	Gender	Selected Response (A)	(B)	(C)	(D)	(E)	(Omitted)
1	F	2.8	13.5	42.2	4.7	25.1	11.6
	M	2.2	13.7	51.6	2.9	20.2	9.3
2	F	11.0	19.3	10.9	6.3	23.8	28.7
	M	7.9	16.7	9.5	6.6	33.2	26.1
3	F	7.7	2.0	58.9	5.0	2.2	24.0
	M	4.5	1.6	74.7	3.5	1.4	14.5
4	F	12.3	4.3	47.4	9.1	3.8	23.2
	M	7.3	3.5	60.1	8.2	3.4	17.4
5	F	10.0	35.2	6.3	7.1	3.5	37.8
	M	9.2	46.4	5.6	5.4	2.7	30.6

N = 99997 (92984 reporting GPA) F = 54606 (51242 with GPA)
M = 45391 (41742 with GPA)

The selected items were presented to the students on five separate pages. Students were expected to enter all written work on these pages. These pages were collected and analyzed.

While each student worked on the problems, the interviewers were instructed to record the student's performance using a coding sheet developed for this study (see Appendix F). This sheet allowed the observer to develop a transcript of the student's performance. This transcript was used during the interview to query the students about their work.

After the students finished working they described how they had worked each problem. The interviewers wrote down what the students said about each problem on the forms provided (see Appendix G). The conversations were also

recorded and transcribed.

Interviewer Training

Four interviewers conducted the student interviews. These persons were all female doctoral students with math or science teaching experience.

The interviewers were trained during one two-hour session, in which each of the five SAT problems were examined and possible solution processes were discussed. The coding sheet was reviewed to insure that each interviewer knew what each symbol meant and interpreted them similarly. The protocol for the study was explained.

Specifically, the interviewers recorded all of the students work behaviors, using the coding sheet, while the students solved each problem. The interviewers were also told to record the time the students spent on each of the questions. When the students finished solving the problems the interviewers asked permission to tape record the next part of the session. The students were then asked to explain their work. While asking the students about their solution processes, the interviewers were instructed to use their codings to prompt students. All of the students' explanations were written down and the sessions were recorded.

Procedure

Student were read the permission/release statement and the introductory statement, and were then asked to allow the session to be recorded. The students were told to solve each problem as they normally would if they encountered this question on the SAT. While the students solved the problems, an interviewer watched and coded the students' performances. When the students finished all of the problems, the interviewer asked the students if they wanted to go back over any of the problems. The amount of time the students spent on each question was recorded.

When the student had completed work on the 5 problems, the interviewers asked the students to explain how they solved the problems, one at a time. Interviewers used the process codings, recorded while the students worked, to questions students about what they had done. The interviewers prompted for student's choice of process, answer selection, and explanation of written work.

After each question was explained, the students were asked to rate the problems. The students rated the items for level of difficulty (very easy, easy, not sure, difficult, and very difficult) both as a math problem and as an SAT problem (see Appendix G).

At the end of the interview, the students were asked to rate how difficult they found the verbal and math PSAT (see Appendix G).

After this, each of the students were told the purpose of the study, which of the problems s/he solved correctly and how the problems s/he didn't solve correctly could have been solved.

Results

It was hypothesized that the percentage of females and males answering each SAT item correctly would be similar to that reported by ETS (see Table 2). It was also hypothesized that the percentage of responses for each alternative would be similar to the actual test populations'. These hypotheses were not supported. The percentages of students were not similar to ETS's statistics (see Tables 4 and 5).

Table 4

Percentage of students answering each item correctly, by gender and race.

Item No.	Subject	Gender	All	White	Black
1	Algebra	F	55	80	30
		M	40	40	40
2	Arithmetic	F	25	50	0
		M	15	10	20
3	Geometry	F	75	100	50
		M	90	90	90
4	Geometry	F	85	100	70
		M	60	50	70
5	Arithmetic	F	65	70	60
		M	55	60	50

Note: N = 40 (10 white females, 10 black females, 10 white males, and 10 black males)

Table 5

Percentage of females and males responding to each alternative, including omits.

Item No.	Gender	Selected Response					
		(A)	(B)	(C)	(D)	(E)	(Omitted)
1	F	0	10	55	5	30	0
	M	0	45	40	5	10	0
2	F	10	40	10	10	25	5
	M	0	35	10	5	15	25
3	F	20	5	75	0	0	0
	M	5	0	90	5	0	0
4	F	15	0	85	0	0	0
	M	20	10	60	5	5	0
5	F	5	65	5	10	0	15
	M	10	55	15	5	0	15

N = 40 (20 females and 20 males)

It was further hypothesized that the time spent on each problem by the females and males would differ (see Table 6). Students' difficulty ratings (see Table 7) and the procedures employed (see Table 8) were also expected to differ for females and males. These hypotheses were supported to some degree.

Table 6

Average time spent answering each item, by gender and race.

Item No.	Subject	Gender	All	White	Black
1	Algebra	F	1.8	1.4	2.3
		M	1.7	1.7	2.6
2	Arithmetic	F	3.1	2.4	3.8
		M	3.9	3.5	4.2
3	Geometry	F	2.8	1.4	4.1
		M	1.5	1.7	1.3
4	Geometry	F	2.4	1.3	3.4
		M	2.3	1.9	2.6
5	Arithmetic	F	3.9	2.9	4.8
		M	2.2	1.9	2.5

Note: N = 40 (20 females, 20 males)

Table 7

Percentage of students responding to each of the difficulty ratings for the five SAT problems and for their PSAT verbal and math performance.

Group	Rating	As Math Problem 1	2	3	4	5	As SAT Problem 1	2	3	4	5	VPSAT	MPSAT
	very easy	0	0	20	10	0	10	0	10	0	0	0	0
White	easy	70	0	10	80	20	70	0	20	70	20	10	20
Female	not sure	30	10	30	0	10	10	40	20	20	20	30	30
	difficult	0	60	20	10	40	10	50	30	10	30	50	50
	very difficult	0	30	20	0	30	0	10	20	0	30	10	0
	very easy	10	0	0	20	0	30	10	0	20	0	0	0
Black	easy	60	10	30	40	10	30	10	30	30	0	40	0
Female	not sure	10	20	40	20	10	20	30	40	30	30	20	20
	difficult	10	40	20	20	60	10	30	20	20	40	40	60
	very difficult	10	30	10	0	20	10	20	10	0	30	0	20
	very easy	10	0	10	20	10	10	0	20	10	10	0	0
White	easy	30	10	60	20	10	20	10	40	40	10	30	20
Male	not sure	20	40	0	30	20	10	50	0	30	10	10	50
	difficult	40	30	30	30	40	50	30	40	20	60	60	30
	very difficult	0	20	0	0	20	10	10	0	0	10	0	0
	very easy	0	0	20	0	0	10	0	20	0	0	0	10
Black	easy	50	0	40	60	20	40	10	30	40	10	20	30
Male	not sure	40	20	10	20	0	10	20	10	50	10	10	40
	difficult	10	70	30	20	60	40	50	40	10	70	60	20
	very difficult	0	10	0	0	20	0	20	0	0	10	10	0

N = 40 (20 females and 20 males)

Table 8

Procedures used for each of the five questions for each of the student groups.

		WF	BF	WM	BM
Question 1					
Used midpoint formula	Correctly	6	3	3	2
	Incorrectly	1	4	2	4
Substituted	Correctly	2	0	1	2
	Incorrectly	1	2	1	1
Guessed	Correctly	0	0	0	0
	Incorrectly	0	1	2	1
Not Classifiable	Correctly	0	0	0	0
	Incorrectly	0	0	1	0
Question 2					
Found common denominator	Correctly	5	0	0	2
	Incorrectly	2	0	2	2
Made equal to 1 cup	Incorrectly	1	5	3	3
Guessed	Correctly	0	0	1	0
	Incorrectly	2	3	1	1
Not Classifiable	Correctly	0	0	0	0
	Incorrectly	0	1	0	0
Left Blank		0	1	3	2
Question 3					
Used Geometry	Correctly	5	1	2	2
	Incorrectly	0	3	0	0
Saw the ratio	Correctly	2	0	1	1
	Incorrectly	0	0	0	0
Examined Answers	Correctly	3	4	5	2
	Incorrectly	0	0	0	0
Guessed	Correctly	0	0	1	1
	Incorrectly	0	1	0	0
Not Classifiable	Correctly	0	0	0	2
	Incorrectly	0	1	1	1
Question 4					
Used Geometry	Correctly	10	5	5	7
	Incorrectly	0	2	3	1
Guessed	Correctly	0	0	0	0
	Incorrectly	0	1	2	2
Question 5					
Used Volume formula	Correctly	4	2	1	1
	Incorrectly	1	0	0	0
Drew Picture	Correctly	2	2	2	2
	Incorrectly	1	0	1	0
Saw the Pattern (ratio)	Correctly	1	0	2	2
	Incorrectly	0	0	0	0
Guessed	Correctly	0	1	1	0
	Incorrectly	1	2	0	3
Not Classifiable	Correctly	0	1	0	0
	Incorrectly	0	0	2	0
Blank		0	2	1	2

In order to fully characterize the students' performance, a description of the students' responses, processes, and difficulty ratings is reported for the five item individually. Specifically, the percentage of correct responses for each student group, the repertoire of solutions, and the students' characterization of the SAT problems, will be described for each of the items.

Problem 1

Problem 1 asked the student to find the coordinate of the midpoint of a given line segment (see Appendix E). In comparison to the mean of the national test population, a greater percentage of the females, both white and black, and of the black males answered this question correctly (see Table 9). A smaller percentage of the white males answered this problem correctly.

Table 9

Comparison to actual test population, as reported by ETS, of percentage correct for Problem 1.

	Actual Test Population[b]				Present Sample[c]		
	All	White	Black	A+ GPA	All	White	Black
Females	42	44	25	66	55	80	30
Males	52	53	32	79	40	40	40

b - N = 99997

c - N = 40

Solution Methods. This problem could be solved in a number of ways. The midpoint formula could be utilized or general test-taking strategies like substitution could be

used. Some students used a combination of both of these approaches. The percentage of students selecting each alternative is given in Table 10.

Table 10

Percentage of females and males responding to each alternative including omits for Problem 1.

Gender	Race	Selected Response					
		(A)	(B)	(C)	(D)	(E)	(Omitted)
F	W	0	10	80	10	0	0
F	B	0	10	30	0	60	0
M	W	0	40	40	10	10	0
M	B	0	50	40	0	10	0

N = 40 (20 females and 20 males)

Many students began solving this item with a formula for midpoint. Some, but not all, of the students recalled the midpoint formula correctly. The ones that did simply found the average, or added x and x+1 and divided this sum by 2. This was how 60% of the white females, 30% of the black females, 30% of the white males, and 20% of the black males reported solving Problem 1. These students correctly selected answer C. As one student explained:

> "Since they are asking for the midpoint of P and Q, I added the two equations (sic) together and divided by 2. . . . I separated the two numbers and I just put 2x over 2 and 1 over 2 and for 2x over 2 I cancelled out the twos."

Other students employing the midpoint formula to get

their answer, stated the formula incorrectly. Most of these students multiplied the two coordinates rather than adding them, thereby arriving at answer E. Sixty percent of the black females selected answer E. It should be noted, however, these girls attend the same school, which may have had something to do with the similarity of their responses. One student explained:

> "They asked you for the midpoint and midpoint is like you say x times x+1 and work the problem out, and then divide it by 2. And that's how I figured out that E would probably be the answer."

Another mistake made by students using the midpoint concept was taking zero rather than x as an endpoint of the line segment. These students took the length of the line segment to be x+1 and the midpoint to be x+1/2 (answer choice B). Forty percent of the white males, 50% of the black males, and 10% of the females chose answer B for Problem 1.

> "They asked what is the coordinate of the midpoint. I figured to get that x+1 and divide by 2. That's why I chose B."

Another reason students gave for choosing B was the appearance of the correct operations - addition and division by 2. This was the rationale used by some of the students

1. Students who reported guessing most often chose B. This was illustrated by the following student's explanation:

> "Well, midpoint of a segment you know is always addition. Then you got to divide it by 2 and what I did was they gave me P and Q so if P is x and Q is x+1, the midpoint of it would be actually what I chose. You gotta divide by 2, x+1 over 2. No. Yes. Actually, looking at it again, I know for a midpoint you must divide by 2 - and that's basically the thought process I used as far as what I know about midpoint and what I have to do with the number I get as a midpoint which I have to divide by 2."

Some students, who got the right answer, did not use the midpoint formula to answer this question, but rather used substitution. Almost one-third of the students (see Table 7) solved Problem 1 by substituting numbers for x and x+1. This method allowed students to simplify the problem by eliminating the variables. For example, as one student explained:

> "I knew the answer had to be between x and x+1. So I substituted a number for x. Then I figured it out. I used each of these answers to get the answers."

This substitution strategy was also sometimes used in order to guess. For example:

"I substituted a number for x. I made x one so that would make Q two. . . . and then I looked at my choices and I saw that x over 2 would put it before P. . . . and B will bring me back to 1 . . . and D and E it would increase instead of decrease . . . .so I guess it would be C."

Substitution was the strategy most often used to arrive at an answer when the math knowledge (ie., midpoint formula) was not known or could not be immediately recalled.

Difficulty Ratings. The students were asked to rate the level of difficulty of Problem 1 both as a math problem and as an SAT problem. Generally, the ratings corresponded with the percentage of students who got the problem right (see Table 7). In other words, the females rated the question easy and the males rated the problem as difficult. However, two of the ten black females rated the problem easier as an SAT problem than as a math problem. Conversely, the black males rated the problem more difficult as an SAT problem. No definite reasons for this shift could be discerned from the transcripts of the interviews.

Time. The time spent on Problem 1 ranged from 10 seconds to 9.5 minutes. The black students spent more time, on average, than the white students. Twelve of the students, five white females, three white males, and four black males,

completed the problem correctly in an average of 2.3 minutes. It should be noted that the amount of time spent on this problem would have negatively affected many students' scores on the actual SAT, since the students only have a total time of one hour to complete 60 math problems.

Problem 2

Problem 2 asked the students to find how many teaspoons of salt would be required if whole cups of both oats and water were used in a recipe for oatmeal cookies (see Appendix A). Comparing the present sample to the national test population shows that about the same percentage of females got the problem right but a smaller percentage of males got it right, if we compare only gender. However, Table 11 also shows that, for the present sample, the white females did better than both the test population and the rest of the present sample. Fifty percent of the white females got the correct answer. In comparison, only one white male, two black males, and none of the black females got Problem 2 right. Further, only one female (black) left the question blank, as compared to three white males and two black males.

Table 11

Comparison to actual test population, as reported by ETS, of percentage correct for Problem 2.

	Actual Test Population[b]				Present Sample[c]		
	All	White	Black	A+ GPA	All	White	Black
Females	24	25	13	45	25	50	0
Males	33	35	17	62	15	10	20

b - N = 99997

c - N = 40

Solution Methods. The key to solving this problem was realizing that the ingredients had to be increased proportionally to get whole amounts of water and oats. The students who got this problem right worked with this concept in some form as illustrated below. Some students mistakenly interpreted the problem, and others guessed. The percentage of students selecting each alternative is denoted in Table 12.

Table 12

Percentage of females and males responding to each alternative including omits for Problem 2.

	Selected Response					
	(A)	(B)	(C)	(D)	(E)	(Omitted)
White Females	10	30	10	0	50	0
Black Females	10	50	10	20	0	10
White Males	20	40	0	0	10	30
Black Males	0	30	20	10	20	20

N = 40 (20 females and 20 males)

Students who got this problem right, answer choice E, worked with the common denominator, 12, to make the

ingredients whole cups. One student explained:

> "...Well first I wanted to see what the relationship between them, in what proportions. . .They wanted how many whole cups you can have and then how much of a teaspoon. Like you increase both by 12, what whole number you use and then still end up with a whole number. . . And then I did 3/4 times 12 and 1/4 by 12. I increased the water by 12 and the oats by 12. . . Then I multiplied the salt by 12 and got three."

Some students began the problem correctly, using the common denominator of 12 and/or converting the fractions to twelfths. However, many of these students then failed to comprehend the idea of increasing the ingredients proportionally. For example:

> "They ask for the least possible multiple of each substance, like in the whole number of cups. So the least multiple of 3/4, 1/4, and 1/3 is in twelfths, which is 9/12, 4/12, and 3/12. They want the recipe to be in whole numbers. So the closest to the whole number is 3/12 away from the one, and the oats is 8/12 from the whole. I figured that 3/12 is .1/4.
>
> Let me get back to the beginning. At 9/12, 3/4 is equal to 9/12. One minus 9/12 is 3/12 which is 1/4 from one and the other one is 8/12 away from one which is 2/4 which I got 3/4. And that's how I got my answer. I

can't really explain. I wasn't really sure about this one."

The idea of making the ingredients equal to one cup was a common misconception. Many of the students thought that the answer to this problem was 3/4, choice B. Many of these students did not see the complexity of the question. Instead they simply ignored the other ingredients and worked only with the amount of salt. This was illustrated by one student who said:

"Well, they told me it's a whole number and they have cups of both water and oats. They want to know how many teaspoons of salt I would require. They give me one quarter teaspoon. They give me 3/4 cup of water and 1/3 cups of oats. I thought to myself, they want a whole number. So salt they give me 1/4. One-fourth plus what will give me a whole number. So I figure 1/4 plus 3/4 which is 4/4 equals 1. That was my thought process."

Similarly another student explained:

". . .I thought like whole numbers like 1 or 2. So I picked 1 and I used the number 1. So from that, they wanted to know how much salt, how many teaspoons were in the salt. . . So I just did 3/4 plus what equals 1? . . . I was thinking about it and the salt it says 1/4 so I said what else, like how much more would give me one. So I just put 3/4."

The other answer choices, A and C were chosen by a few students. These students admitted they were basically guessing. For example:

> "This one was pretty much a guess . . . since there's 3/4 cup of water and 1/3 cup of oats . . . you'd need about half a teaspoon to make it even. So I put A, a half."

Another student explained:

> "I always have trouble with fractions. . . .They wanted the number of cups to be whole so I was going through the numbers trying to make it whole. . .Then I realized they wanted the whole number of salt and the whole number of water and oats. So I tried using these to make a whole number of water. . . And at the same time after I made a whole I made the same ones to find out how many teaspoons were in it. . . It didn't work so I guessed. I just figured 1 teaspoon."

Difficulty Ratings. As with the other problems, the students were asked to rate the level of difficulty of Problem 2 both as a math problem and as an SAT problem. Generally the students rated this problem as difficult, although some groups were more definite than others (see Table 7). Specifically, 90% of the white females, 70% of the black females, and 80% of the black males rated the problem

either difficult or very difficult as a math problem. Only 50% of the males rated the problem difficult or very difficult; 40 were not sure.

Some of the students were less sure of the difficulty of Problem 2 as an SAT problem. Sixty percent of the white females, 50% of the black females, 40% of the white males, and 70% of the black males rated the problem as either difficult or very difficult. Forty percent of the white females, 30% of the black females, 50% of the white males, and 20% of the black males were not sure. This may have to do with an opinion some students held that there is a shortcut way to solve most SAT problems. This was verbalized best by one student who said:

> "This shouldn't have been so hard. There's always a way you can get it. I mean, for the SAT they always sneak in a way. There's always a trick, a shortcut but I couldn't get it. It was too hard."

Time. The amount of time the students spent working on Problem 2 ranged from one minute to 11 minutes. The average amount of time spent was 3.5 minutes (see Table 6). Again it should be noted that most of the students spent much more time on this problem than they would have been able to in order to complete the actual test. Only six out of the 40 students answered this question in one minute, but none of these six got the right answer. The average time spent by

those students who got the right answer was 2.9 minutes.

Problem 3

Problem 3 involved the geometric concepts of length, diameter, and tangent (see Appendix E). The majority of the students answered this question correctly (see Table 13). More students got the right answer than expected based on ETS's percentages. All of the white females, 50% of the black females, and 90% of the males got the correct answer (see Table 13). However, as illustrated below, many of the students did not use the geometric concepts to get the answer. They simply examined the answers and found the one that was unlike the others. The percentage of students selecting each answer option (see Table 14), and their solution methods are described below.

Table 13

Comparison to actual test population, as reported by ETS, of percentage correct for Problem 3.

	Actual Test Population[b]				Present Sample[c]		
	All	White	Black	A+ GPA	All	White	Black
Females	58	61	38	77	75	100	50
Males	74	77	51	90	90	90	90

b - N = 99997

c - N = 40

Table 14

Percentage of females and males responding to each alternative including omits for Problem 3.

	Selected Response					
	(A)	(B)	(C)	(D)	(E)	(Omitted)
White Females	0	0	100	0	0	0
Black Females	40	10	50	0	0	0
White Males	10	0	90	0	0	0
Black Males	0	0	90	10	0	0

N = 40 (20 females and 20 males)

Solution Methods. The diameters of the circles were equal to the width of the rectangle. The length was equal to twice the diameter. Therefore, the length of the rectangle was twice the width of the rectangle. This meant that the answer choice that did not contain numbers in a two to one ratio, answer C, was the correct choice (the problem asked which CANNOT be the length and width).

As explained above, the key to solving Problem 3 involved seeing the two to one ratio of the length and width of the rectangle.

Some students, especially the white females, took an approach to this problem that involved establishing the geometric relationship and then finding the appropriate answer. For example:

> "We're given a rectangle with two circles inside and we know that the diameter of one circle is half the length of the rectangle. Since you have two circles, the length of the rectangle would equal twice the diameter

of one of the circles and the width of the rectangle would be equal to the diameter of the circles which would be the same. You were asked which of the pairs could not be the length and the width. So it would be in a proportion of two to one, length to width. . . . and the only one there, not in a two to one ratio, is 16 to 10."

Many students, especially the males, who got this problem right did not even consider the geometry. Instead they just looked at the answers and picked the answer choice that did not relate to the other ones. As one student explained:

"Well I read the question, looked at the picture and I looked at the choices and I realized there was a pattern. Well if you took the second number and multiplied by two you would get the first number, except for choice C."

Choice A was the distractor chosen by four of the 10 black females, and one black male. Their explanations indicated that some misread of the question, missing the "CANNOT". For example:

"I put A cause I figured the length was half the width, so I put two to one."

Others guessed, or answered based on the way the picture

looked. As one student explained:

> ". . .It's in the figure. I don't picture a rectangle with the largest side being two and then the smaller side being one."

Difficulty Ratings. The males were more definite about their ratings for this problem than the females (see Table 7). Seventy percent of the white males and 60% of the black males rated this problem easy or very easy. The remaining males rated the problem as difficult. The boys rated the problem similarly as a math problem and as an SAT problem.

The females were not as definite. Thirty percent of the white females rated the problem easy or very easy, 40% rated the problem difficult or very difficult, and 30% were not sure. The black females' ratings were similar. Like the boys, the girls' rating did not change when viewing the problem as a math problem or as an SAT problem.

Time. The time students spent working on Problem 3 ranged from 10 seconds to 9.5 minutes (see Table 6). The average time was 2.15 minutes. The black females spent more time on Problem 3 than the other groups, an average of 4.1 minutes. The 33 students who answered Problem 3 correctly spent an average of 1.9 minutes on this problem.

Problem 4

Problem 4 asked the student to find a value for one of the angles, marked x, which was contained in the picture accompanying the problem (see Appendix E). All of the white females, 70% of the black females, 70% of the black males, and 50% of the white males solved this problem correctly (see Table 15). Not one student left this problem blank, ie., omitted the problem. In comparison to the actual test population, the percentages were very different than expected. The specific solution methods employed, and the answers chosen are discussed below (see Table 16).

Table 15

Comparison to actual test population, as reported by ETS, of percentage correct for Problem 4.

	Actual Test Population[b]				Present Sample[c]		
	All	White	Black	A+ GPA	All	White	Black
Females	47	50	25	72	85	100	70
Males	60	62	34	85	60	50	70

b - N = 99997

c - N = 40

Table 16

Percentage of females and males responding to each alternative including omits for Problem 4.

	Selected Response					
	(A)	(B)	(C)	(D)	(E)	(Omitted)
White Females	0	0	100	0	0	0
Black Females	30	0	70	0	0	0
White Males	30	10	50	10	0	0
Black Males	10	10	70	0	10	0

N = 40 (20 females and 20 males)

Solution Methods. To solve this problem the students needed to know some geometric principles related to angles and triangles. Specifically, students needed to know that the sum of the angles of any triangle is 180 degrees and that supplementary angles add up to 180 degrees. Further, students had to recognize that a right triangle with two equal sides has one 90 degree angle and two 45 degree angles. It should be noted that ETS provides the information about the sum of the angles of a triangle on the top of the SAT. However, the students participating in this study were not given this information.

Based on the information given in the problem and the principles elucidated above, the students could find the value of x by deriving the values for angles RSP, PSQ, QST, RPS, and finally SPQ. This involved seeing that the right angle, which was divided into three equal parts formed three 30 degree angles (<RSP, <PSQ, and <QST), and that RPS must equal 105 degrees. The students could then arrive at the value of 75 degrees for <SPQ; answer choice C. The majority of the students who got this problem right solved it this way.

The students' explanations of how they solved Problem 4 indicated that many of the students interviewed understood the geometric concepts described above. For example, one student explains:

"Well, they said that this was 90 degrees so, and the

three triangles are equal so it's 30 degrees each angle. This is equal, so that's isosceles. So if this is 30, this is 45, and the whole triangle is 180, then this has to be 105. And if this is supplementary, this is x so this has to be 75 and it's x."

No other method of getting the right answer was offered by any of the students. However, some students, who got the wrong answer, offered other methods. For the most part, the students' explanations for arriving at the other answer choices were seemingly based on their visual perceptions of the diagram given. For example:

"I just drew lines through the picture to find what the right angle R is. We had to find the angle of SPQ, so I just . . .the line RT was 180 degrees. So I just drew a line and estimated the amount it would be, and it seemed a little bit less than 90 and more than 80. So I took 85."

The most common wrong answer choice given was A, 65. Students seemingly picked this answer because of its closeness to what looked right. One student explained:

"They told me the triangle was a right triangle, and if RS is congruent to ST, the right angle RST is divided into three equal angles. So if they're equal, they have to be the same. So if RST, with S right here, what I

did was put it all into equal segments, 30, 30, 30, which is 90. And I thought of a way to get x become a complimentary of some sort if they're all equal. Because to me, if you look at the diagram, it looks like an equilateral. They're not telling me that RP is congruent to QT but it looks that way. X seems to be kind of complementary. . .I tried to figure out what number will give me a complementary angle x to basically be given equal sides of S and that's how I went about getting a number. . .I looked at it as a 30-60-90 triangle. . .And I tried to pick something to be closest to what I assumed to be the answer."

Difficulty Ratings. The students' ratings paralleled their level of success. Ninety percent of the white females rated this problem as easy or very easy as a math problem and 70% rated it easy as an SAT problem. Sixty percent of the black females rated Problem 4 easy or very easy as a math problem. Fifty percent of them rated the problem easy or very easy as an SAT problem.

Sixty percent of the black males rated Problem 4 easy or very easy as a math problem. However when asked to rate the problem as an SAT item 40% of the black males rated it easy and 50% were not sure. For the white males, 40% rated the problem easy or very easy as a math problem, 30% rated it difficult, and 30% were not sure. As an SAT problem, the

ratings were similar.

Time. The time for solving Problem 4 ranged from 11 seconds to 6 minutes, with an overall average time of 2.35 minutes (see Table 6). The black students took more time, on average, than the white students. The average time for the students who got the right answer was 1.8 minutes. As with the other problems, this average time exceeds the realistic time students would have to solve this problem during the actual SAT.

Problem 5

Problem 5 asked the students to find how many rectangular pieces a cake could be cut into so that no cake remained (see Appendix E). Many more students than were expected, especially the females, got this problem right (see Table 17). Specifically, 70% of the white females, 60% of the black females, 60% of the white males, and 50% of the black males got the correct answer. Two black females, two black males, one white female and one white male left this question blank (see Table 18).

Table 17
Comparison to actual test population, as reported by ETS, of percentage correct for Problem 5.

	Actual Test Population[b]				Present Sample[c]		
	All	White	Black	A+ GPA	All	White	Black
Females	35	37	20	54	75	70	60
Males	46	49	24	70	55	60	50

b - N = 99997

c - N = 40

Table 18

Percentage of females and males responding to each alternative including omits for Problem 5.

	Selected Response					
	(A)	(B)	(C)	(D)	(E)	(Omitted)
White Females	10	70	10	0	0	10
Black Females	0	60	10	10	0	20
White Males	10	60	20	0	0	10
Black Males	10	50	10	10	0	20

N = 40 (20 females and 20 males)

Solution Methods. This problem could be solved either by using the volume formula or by drawing the picture and cutting it up into the smaller sections. As will be shown below, the students used both of these methods, as well as a combination of the two.

Two basic methods were employed to get the answer to Problem 5. One involved dividing the volume of the cake by the volume of the smaller piece to find out how many pieces. One student explained it this way:

> ". . .I multiplied the three numbers together since the volume was by multiplying three numbers. I multiplied the other volume they gave me and I divided to see how many times the smaller volume went into the larger volume and got my answer."

In a variation on this method, some students realized they did not have to work with the third dimension (2 inches), since it was the same for both rectangular pieces. For

example:

"At first I multiplied all three numbers and I tried to make them equal to each other but it didn't come out. But then I realized that the two inches didn't have anything to do with it. So the area of the rectangle is length times width. So I multiplied the numbers together, 9 times 13 and 3 times 3 and 1/4, and then I divided 117 by 9.75 and got 12."

Some students started out to find the quotient of the volumes but got tangled up in the arithmetic. For example:

". . .The way I tried to figure it out is you multiply 9, 13, and 2 and you get 234. And also multiply 3 by 3 1/4 by 2 and you get 19.5 cubic inches. And I assumed that you can just divide the 234 by 19.5 but either I did something wrong that I didn't get one of the answers or I should have changed it. I should have cubed the answer first, I mean the 19.5 cubed and also the 234 cubed and multiplied that 3 times by itself. Divide and find the cube root, then maybe I would have gotten the answer."

The other method of solving Problem 5 involved drawing the picture and dividing it into pieces. One student explained:

"Since they didn't draw a picture, first I drew one

because I wanted to get a view of what I had to do. And then I worked with the one that says 3 1/4 because since it was divided into parts that were three inches by 3 1/4 inches by 2, I knew that 2 would correspond to the figures and that the 3 would correspond to the 9. So I just had to see how 3 1/4 would go to 13. So I just said well okay, four would make one. So I saw how many. Four threes equals 12, so I knew that that would be the 13 and then I just drew the divisions and I saw that it was 12.

Difficulty Ratings. Most of the students rated this problem as either difficult or very difficult, both as a math problem and as an SAT problem (see Table 7). Sixty percent of the white females, 80% of the black females, 60% of the white males, and 80% of the black males rated Problem 5 as difficult or very difficult as a math problem. Seventy percent of the white females, 70% of the black females, 70% of the white males, and 80% of the black males rated Problem 5 as difficult or very difficult as an SAT problem.

Time. The average time spent on Problem 5 was 3.1 minutes (see Table 6). The times ranged from 1 minute to 10 minutes. The females spent more time, on average, on this problem than the males. The students who answered this problem correctly spent an average of 3.3 minutes, far more

time then would be realistic to spend during the actual test.

Discussion

This study was an initial effort to understand why some groups of students, with essentially the same math knowledge, outperform others on math SAT questions. Based on past research (Bond, 1988; Marshall & Smith, 1987), it was hypothesized that the methods students employ to solve SAT math problems may have some influence on their final score. Specifically, this study focused on the differences in the problem solving methods use by females and males, black and white.

The Educational Testing Service (ETS) provides statistics regarding the percentage of students answering each question correctly by gender and race. These statistics were used to choose the five problems to be included in this study. The chosen problems were the ones with the greatest gender differences for the students with the highest grade point averages. It was expected that about the same proportion of students participating in this study would get the problems correct. This was not the case. More students got the problems right during this study than we had expected, based on nationally reported test statistics.

The fact that many of the students involved in this study were able to correctly solve the problems suggests that

factors other than math ability may be contributing to lowered performance. One possible reason for this may be the time factor. Unlike the actual test, the study was conducted without time restraints. Students were told to solve the problems as they would on the SAT but no time limit was imposed. Many students spent a considerable amount of time solving the problems, as much as 29 minutes. The average overall time spent solving the five problems was 12.9 minutes. Spending this amount of time on five problems would result in a very low final score.

Students did not have to constantly monitor the amount of time they were spending on each problem. Nor did they have to worry that the time spent on one problem would have to be taken from another one. Some of the students commented on the time factor when talking about the PSAT. For example:

> ". . . I think the biggest problem was the time limit. I couldn't, I thought that I had less than a minute for each one because there was a lot of problems and so little time and I just wanted to get through."

Another student, when asked why she rated the problem harder as an SAT problem stated "you have time on the SAT, there are limited time periods."

The black females, who had lower PSAT scores than the other groups, took the most time, on average. Having the time to spend on each problem may have resulted in getting a

greater percentage of the problems correct. If this is true than having to work within a limited time frame (60 minutes) may be negatively affecting their SAT scores.

However, time spent did not guarantee the correct solution. Some of the students, especially two of the black females, spent considerable time working on problems even when they clearly did not know how to solve them. These students persisted beyond a reasonable time limit instead of omitting the question or coming back to it after the other problems were attempted. Problems answered incorrectly lower students' final score; items omitted do not.

This behavior may not generalize to the testing situation but it indicates a possible lack of test-taking skills. When asked why she sat staring at one problem for almost five minutes, one student said:

> "I just look at it 'til I remember. I wait for the formula to like pop into my head, but nothing happened."

Looking at the ways that the students solved the various problems provided some insights into the thought processes of the females and males. For this small sample, the females tended to rely more on their math knowledge than on their test-taking skills.

Problem 3 illustrated these variations. The boys, more often than the girls, said they just looked at the answers and picked the one that didn't fit. The girls, on the other

hand, more often stated the relationship in terms of the length and the width of the rectangle and their relation to the diameters. Similarly, for Problem 5, the females more often used the volume formula and performed the division 234 by 19.5. In contrast, the males explained their answers in terms of the picture and cutting the cake up into pieces.

These differences may suggest that the females, more often than the males, tried to apply what they had learned in math class. Math lessons often involve learning a specific procedure that can be utilized to solve a certain problem type. One study of elementary school students has shown that girls follow these established procedures more often than boys (Grieb & Easley, 1984). The higher grades that they earn in school may be further testimony to this idea - girls do what they are taught in school.

Performing well on the SAT math requires more than just math knowledge. Test-taking skills are important as well. Some of the students acknowledged this, especially the ones enrolled in test-preparation. As one student explained:

> ". . .they have lots of tricks. I mean, I know they can't teach me anything, like you know, that my teachers have taught me, but they teach me ways around it. Like looking at the answers and seeing, like you know, what answer looks different from the rest, or something. That's what they say you should do for the hard questions. When you have absolutely no idea, you should

see how they differ and then the one that is different is the answer. . ."

Even though, ironically, this student is looking for another set of procedures to follow, she expressed the notion that getting a good score is more than just knowing math.

This students' comments also highlight the focus schools put on math. Math is often taught as a sequence of problems, with one right answer, that can be solved using a series of rules, algorithms, and procedures. Tests like the SAT reenforce this view of math. If high SAT scores are important, than math teachers, according to the present model, may have to expand the set of rules and procedures taught in math class. Test-taking skills may improve students' mathematical metacognition. However, emphasizing standardized test scores, like the SAT, may lead math curricula to become more specialized; geared towards standardized test problems, not realistic problem solving. Instead of integrating math into students' lives, math learning will continue to be focused on getting the right answer - quickly.

Suggestions for Future Research

This study highlighted the importance of studying the reasons for students' varying SAT performance. No other studies have examined how students actually solved SAT math problems. Marshall & Smith (1987) characterized the

computation and problem solving performance of third and sixth graders on the California Assessment Program. They found that girls and boys make different kinds of errors. Bond (1988) found that knowledge of math was not enough to achieve a high SAT math score; procedural skills are needed. Further research of students' solution methods, based on the findings of the present study, needs to be done. An evaluation of the mathematical procedures students employ when solving SAT questions should to be done with a representative student sample.

Students' math skills need to be evaluated in more than one way. Research has shown that students with equivalent grades do not perform similarly on the SAT (Rosser, 1989). If one were to judge the math skills of the students involved in this study by their present performance on these five SAT question, the white males would not have outperformed the other students. This is not the reality of the SAT.

Future research needs to focus on the time factor affiliated with the SAT. Would females' performance improve if they had unlimited time? Does the time limitation hinder (or help) some groups more than others, and if so, is this disadvantage justified in terms of predicting first year college grades? A specific study could be done to determine how the time factor influences SAT math scores. Taking away the time restraint would minimize the competitive nature of the SAT. Every student would be given the chance to do her

or his best. Students' timed and untimed performances could be compared.

Another study could be done to examine the impact of the penalty for wrong answers. Students lose a quarter or a third of a point for each problem that they get wrong, whereas there is no penalty for omitting a test item. Looking at the difference in the rate of wrong answers among different gender and ethnic groups may indicate whether or not the penalty for wrong answers should be eliminated. This has already been done for the Graduate Record Exam. These kinds of studies could help identify what factors, other than math aptitude, influence students' SAT scores.

DO FEMALES AND MALES HAVE DIFFERENT PROBLEM SOLVING STYLES THAT CREATE OR CONTRIBUTE TO THE SAT GENDER GAP?

An Analysis with Policy Recommendations by Phyllis Rosser

The sample population for this study was extremely small - 20 students were interviewed after answering eight Verbal questions and 40 students were interviewed after answering five Math questions - so the findings can not be generalized to a larger test taking population. However, the small sample size allowed us to do a fine-grained analysis of individual differences in problem-solving strategies. We found no significant differences in intellectual problem-solving or cognitive processing but did find several important differences in young women's test-taking strategies compared to young men's. We also found no problem-solving differences between African American and white students except in the analogy section of the Verbal test. Our findings were consistent with other research findings and gave strong indications of areas where further research would yield valuable information.

PERFORMANCE COMPARISONS

Ross found that the ten males in her sample performed better than the ten females in her analysis of students from Seattle public high schools. Males answered correctly 38 percent of the time (30 times out of a possible 80) while females answered 30 percent of the questions correctly (24 times).

In Kelly-Benjamin's sample of 20 females and 20 males from New York City's public, private and parochial schools, white females answered the greatest number of questions correctly. They were followed by African American males, white males and African American females. This was one of the most important findings of the study - that females did the best on questions that had been identified as extremely difficult for a national sample of high performing

males and females. Kelly-Benjamin's findings suggest that when females are given enough time to work the SAT math problems, they may have more correct responses than males, suggesting that their lower SAT Math score averages may not be entirely attributable to deficiencies in math knowledge as ETS has claimed. Females answered correctly 61 percent of the time (61 times out of a possible 100 opportunities) while males answered 52 percent of the questions correctly. White females answered correctly 40 of the 61 times while African American females answered correctly 21 times. White males answered correctly 25 times and African American males, 27 times.

DIFFERENCES IN PROBLEM SOLVING STYLES

Ross found there were no major differences between the sexes in problem-solving techniques used to obtain both correct and incorrect answers on the Verbal test. Students of both sexes who answered correctly were similar in analytical skills and knowledge of vocabulary. These students knew more of the required vocabulary, had a more systematic way of making decisions, and had strategic approaches to finding the correct answers (changing strategies for each type of question). Only four of her students had attended SAT coaching courses so the strategies were largely self-developed.

However, African American students in Ross's research were more likely to make vertical comparisons in analogy questions, matching words on each side of the pair instead of matching the horizontal relationship of the paired words, as required. This supported earlier findings by ETS researchers Schmitt and Bleistein (1987) who suggested that African American and white students may be differentially familiar with analogical reasoning.

Kelly-Benjamin found no cognitive differences between the sexes in math problem solving but there was a difference in the use of test taking strategies. Although both sexes were familiar with the same types of

strategies, females tended to rely more on their math knowledge to solve the math problems. White females were more likely to employ the math methods they had learned in the classroom - using formulas or knowledge of geometry rather than their test taking skills. This supports earlier research by Grieb & Easley (1984) who found that elementary school girls followed established math procedures more often than boys. White males were more likely to subsitute answer choices until they found the correct one or use other test-taking strategies.

All the white females answered the two geometry problems correctly, despite research showing that geometry is differentially harder for women (Donlon, 1971; McPeek and Wild, 1987; Rosser, 1989). Sixty percent of the African American girls chose the same incorrect distractor for Question 1. Kelly-Benjamin noted that this may have resulted from the use of an incorrect algorithm; these students attended the same school.

DIFFICULTY RATINGS BY STUDENTS

Ross reports that both sexes found each of the Verbal questions "difficult" or "very difficult" in about the same degree with the exception of Question 3 about sports journalism. This question turned out to be easier for males and was answered correctly by more African American males than white males. The question rated most difficult by both sexes and races was Question 5 (Dessicate: Moisture), primarily because none of the students were familiar with the word "Dessicate." However, it was answered correctly by three females (two white, one African American) and two males (one white, one African American), who deciphered the meaning of the word using horizontal comparisons.

Kelly-Benjamin found Math Question 1 (using algebra to find the midpoint of a line) was rated easier by women than men (none of the white women and only

two of the African American women thought it was a difficult math problem). On the other hand, Question 2 (an oatmeal recipe requiring only arithmetic knowledge) was rated somewhat easier by men. However, a larger percentage of white females answered both questions correctly. The men also found Question 4 (using geometry to find the value of an angle) somewhat harder than the women did. In fact, white women, African American women and African American men all outperformed white men on this question.

TIME PRESSURE

One of this study's most important findings was that eliminating the speeded nature of the SAT improved female performance on both the Verbal and the Math questions. In fact, when allowed unlimited time, white females performed better than white males in Math and equally well in the Verbal sections. Earlier research by Rosser (1989) and Graf and Riddell (1972) found that girls moved more slowly through the Math sections. This study investigated this more closely by giving students all the time they required to answer both verbal and math questions rather than the 40 - 60 second average allowed on the actual SAT. The average number of seconds each student used to answer each question indicated that most students took more than 60 seconds to interpret and answer the items.

Ross found an interesting discrepency between the sexes. Correct female responders took more time per question than the female sample in general and more than the average allotted time (40 seconds) on six of the eight questions. However, correct male responders took less time per question than the males in general and more than 40 seconds on only four of the eight questions. Since both sexes were approximately equal in knowledge of vocabulary and use of analytical strategies, this suggests that increasing the time has a beneficial

effect on female performance but not on male performance.

Kelly-Benjamin found that correct responders of both sexes averaged more than the allotted SAT time on every question, but white females - who had the highest number of correct answers - averaged the least amount of time per question. Increasing the allotted time did not seem to help the boys since they spent more time, on average, than the white females and answered fewer questions correctly. Perhaps not having to worry about time influenced the girls' approach to the problems but not the boys.

This is evidenced by the solution methods the students employed. White girls applied the mathematical procedures they had learned in the math classroom more often, whereas boys used test taking skills more frequently. Since classroom math usually involves a set series of procedures, this suggests that extra time benefits girls by allowing them to perform math operations they wouldn't have had time to do in the 60 second average allotted for actual test questions.

Ross found in her racial analysis that white females averaged the least amount of time per question on the Verbal test and African American males averaged the most. However, speed did not correlate positively with correct responses, as noted above. Kelly-Benjamin found that white students averaged less time per question (except for Question 3 where African American males averaged the least time) than African Americans of both sexes. African American girls took the most time, on average, and answered the fewest number of questions correctly. Two African American girls persisted beyond a reasonable time limit, even in the extended time situation, indicating a lack of test-taking skills.

OMISSION OF QUESTIONS

Our findings here raise important questions although we were unable to compare Verbal and Math students directly because of a difference in procedures: Ross encouraged her students to answer each question and had only one (African American male) refuse to guess. However, she did ask her students whether they would have omitted the question on an actual SAT. Kelly-Benjamin gave each student the option of omitting during the interview process but did not ask them what they would have done on the actual test.

Ross' research suggests that girls are less likely to guess than boys but when they do guess, they are more likely to choose the right answers. Earlier research also found that, in the science and math areas, girls were less likely to guess than boys, but more likely to guess correctly. Linn, DeBenedictis, Delucchi, Harris and Stage (1987) and Miller and Suchner (Rosser, 1989, p.64) found that girls scored higher, even higher than boys in math, when the "I Don't Know Option" was removed from the NAEP science and math tests. Kelly-Benjamin's findings showed that when females have more time to work the math problems they are less likely to omit them (compared to Rosser's findings on the November 1987 SAT where a larger percentage of girls than boys omitted 50 of the 60 math questions in the speeded environment of the actual test). In other words, girls know the math (confirmed by their higher grade averages) and when they have time to carry out the procedures they have learned in school, the need to omit rather than be penalized for wrong answers is minimized.

Ross found a larger number of females than males would have omitted the Verbal questions on an actual SAT. Girls said they would have left questions blank 27 times out of a possible 80 while boys said they would have left questions blank 15 times. However, when asked to guess, girls answered correctly six times while boys answered correctly only three times. This finding confirms the findings of Linn, et al. (1987) and Miller and Suchner

(Rosser, 1989, p.64).

Furthermore, Ross found that when she questioned students about omitting questions on an actual SAT, six of the 10 boys wanted to know if there was a penalty for quessing while only two of the 10 females did. This indicates a test-wiseness or gaming strategy toward the SAT that may be more prevalent among boys than girls. In fact, Harvard Education Professor Carol Gilligan (personal communication, 1987) said, "the SAT is a moral issue for girls, they think it measures their intelligence and they mustn't cheat (by guessing) but boys rise to the competitive challenge of the test and play it like a pinball game."

Kelly-Benjamin found that when given the time to work the math problems, girls omitted fewer questions than boys in her math sample. In fact, there were omissions on only two questions: on Question 2 (oatmeal recipe) one female and five males left a blank and on Question 5 (rectangular cake) three males and three females left blanks. It is interesting to note that both of these questions involved experiential context supposedly more familiar to girls.

Ross found no significant differences in intended omission between white girls and African American girls except for Question 8 (opposite of Haphazard). There were also no differences in omissions between white males and African American males. Kelly-Benjamin found no omissions by white females. However white and African American males omitted slightly more (four) questions than African American females (three).

All of these findings indicate that risk taking is an area deserving further study and that it may yield useful information in determining why the SAT gender gap persists and widens.

SUMMARY OF FINDINGS - REWARDING THE TESTWISE RATHER THAN THE WISE

Test critics and critical thinking experts agree with Mathematics Professor Kate R. Sheehan who said recently in The New York Times (July 21, 1989) that "speeded or timed tests reward the facile test taker, not the sophisticated, thoughtful thinker who gathers information and organizes, evaluates and expresses ideas clearly, perceives subtle relationships and patterns, and makes reasonable estimates and predictions."

The present findings tend to support this view. Girls who responded correctly on both sections of the test averaged more time per question than high scorers would have used on an actual SAT. Ross found that correctly responding girls took more time than correctly responding boys or than her female sample, in general. However, correctly responding boys took less time than the male sample in general and for half the questions, less than the SAT allotment.

Kelly-Benjamin found that correctly responding girls took more than the allotted SAT average time per question. This worked to the advantage of the white girls who were most likely to perform the step-by-step math operations as they had learned them in the classroom. Boys, on the other hand, more often used the less time consuming strategy of plugging answer choices into the questions until they found the correct one. This strategy was particularly apparent in Question 3 where girls were more likely to establish the geometric relationship while boys found the correct answer by looking at the choices. However, increasing the time was not as beneficial for African American students who, in general, averaged more time per question than the white students. Two of the African American girls persisted beyond a reasonable time limit when they clearly did not know how to solve the problem.

Ross also found that girls were reluctant to take risks and guess when they didn't know the vocabulary words. They reported that they would omit

nearly twice as many questions as boys. However, when forced to guess, they answered twice as many correctly. Kelly-Benjamin found that the number of omitted questions for each group was extremely small, suggesting that when students have enough time to work the math problems, the risk taking factor is removed.

As stated earlier, the sample of students in this study is too small to provide conclusions that can be generalized to a national test taking population. However, the findings do suggest that students who use strategies that promote rapid decision making will receive higher SAT scores than students who have the requisite academic knowledge but are not facile test takers. Since our results show a difference between the sexes and races, we conclude that it would be highly beneficial to conduct more research in these areas.

Speed, risk taking, and the use of test taking strategies are also linked to self confidence, test anxiety and perception of tests as accurate measures of ones ability - psychological factors which were not part of this study. We suggest that further research be done in these areas, by gender and race.

We conclude that the SAT may not be discriminating fairly for success in the first college year because students who have the knowledge and academic skills to perform well are being penalized for their lack of test-taking skills. It is unfortunate and unfair to allocate resources based on testwiseness rather than critical thinking ability or acquired knowledge in subject areas. It also seems a waste of time and money to make SAT test-taking preparation a prerequisite for success. It is our hope that these pilot studies will encourage and inspire other researchers to investigate the areas studied here with larger sample populations. We think it is imperative that the SAT's underprediction of first year college performance for women and minority men be understood and remedied.

BIBLIOGRAPHY AND REFERENCES

ACT. 1987. College Student Profiles: Norms for the ACT Assessment. Iowa City: American College Testing Program.

Bond, Lloyd. 1988. Understanding the black/white student gap on measures of quantitative reasoning. Paper presented at Howard University, Washington, D.C., Testing and the Allocation of Opportunities in Education and Employment to Black Americans. December.

Boyer, Ernest L. 1987. College: The Undergraduate Experience in America. New York: Harper and Row.

Burton, Nancy W., Charles Lewis, and Nancy Robertson. 1988. Sex Differences in SAT Scores. Report No. 88-58. New York: College Entrance Examination Board.

Clark, Mary Jo, and Jerilee Grandy. 1984. Sex Differences in the Academic Performance of Scholastic Aptitude Test Takers. Report No. 84-8. New York: College Entrance Examination Board.

Coffman, William E. 1961. "Sex Differences in Response to Items in an Aptitude Test." National Council on Measurement in Education, Eighteenth Yearbook. Michigan. pp. 117-124.

Donlon, Thomas F. 1971. Content factors in sex differences on test questions. Paper presented at the Educational Research Organization, Chestnut Hill, Massachusetts.

Donlon, Thomas F., and William H. Angoff. 1971. The Scholastic Aptitude Test." The College Board Admissions Testing Program, William H. Angoff, ed., pp. 15-47, New York: College Entrance Examination Board.

Doolittle, Allen E. 1985. Understanding differential item performance as a consequence of gender differences in academic backgrounds. Paper presented at the annual meeting of the American Educational Research Association, 31 March - 4 April, Chicago.

Dwyer, Carol A. 1979. "The Role of Tests and Their Construction in Producing Apparent Sex-Related Differences. in Sex-Related Differences in Cognitive Functioning, Wittig and Peterson, eds., New York: Academic Press.

Dwyer, Carol A. 1976a. "Test Content and Sex Differences in Reading." The Reading Teacher 29(8):753-77.

Dwyer, Carol A. 1976b. Test content in mathematics and science: the consideration of sex. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Feingold, Alan. 1988. "Cognitive Gender Differences Are Disappearing." American Psychologist, February, pp. 95-103.

Gamache, LeAnn M., and Melvin Novick. 1985. "Choice of Variables and Gender

Differentiated Prediction Within Selected Academic Programs." Journal of Educational Measurements 22(1):53-70.

Graf, Richard G. and Jeanne C. Riddell. 1972. "Sex Differences in Problem Solving as a Function of a Problem Context." Journal of Educational Research 65(10):451-2.

Grieb, A. and J. Easley. 1984. "A Primary School Impediment to Math Equity: Case Studies in Rule-Dependent Socialization." in Advances in Motivation and Achievement Vol. 2, pp. 317-362, M. SteinKamp and M. Maehr, eds., New York: JAI Press.

Gross, Susan. 1988. Participation and Performance of Women and Minorities in Mathematics. Maryland: Department of Educational Accountability, Montgomery County Schools. July.

Hyde, Janet Shibley, Elizabeth Fennema, and Susan J. Lamon. 1989. "Gender Differences in Mathematics Performance: A Meta-Analysis. Psychological Bulletin (In Press).

Hyde, Janet Shibley, and Marcia C. Linn. 1988. "Gender Differences in Verbal Ability: A Meta-Analysis." Psychological Bulletin 104(1):53-69.

Kanarek, Ellen. 1988. Gender differences in freshman performance and their relationship to use of the SAT in admissions. Paper presented at the annual meeting of the National Council of Measurement in Education, 1-4 April, Chicago.

Lockheed, Marlaine E. 1973. Sex bias in educational testing: A sociologist's perspective. Paper presented at the International Symposium on Educational Testing, The Hague, July.

Linn, Marcia C., Tina DeBenedictis, Kevin Delucchi, Abigail Harris and Elizabeth Stage. 1987. "Gender Differences in National Assessment of Educational Progress Science Items: What Does 'I Don't Know Really Mean?" Journal of Research in Science Teaching Vol. 24, No. 3, pp. 267-278.

Linn, Marcia C. 1985. "Gender Equity in Computer Learning Environments." Computers and the Social Sciences Vol. 1, No. 1, pp. 19-27.

Lubetkin, Julie. 1988. The Scholastic Aptitude Test: A Valid and Unbiased Predictor of College Performance?" Bachelor of Arts thesis, Princeton University, Princeton, New Jersey.

Marshall, Sandra P., and Julie D. Smith. 1987. "Sex Differences in Learning Mathematics: a Longitudinal Study with Item and Error Analyses." Journal of Educational Psychology, 79, 372-383.

McCarthy, K. 1975. Sex bias in tests of mathematical aptitude. Doctoral dissertation, City University of New York, New York.

McPeek, W. Miles, and Cheryl Wild. 1987. Characteristics of quantitative items that function differently for men and women. Paper presented at the annual meeting of the American Psychological Association, August, New York.

Nairn, Allan. 1980. The Reign of ETS: The Corporation that Makes Up Minds. (Unpublished Book)

National Merit Scholarship Corporation. 1988 Annual Report 1987-88. Evanston, Illinois.

Rigol, Gretchen W. 1989. "Simplistic and Unhelpful to Blame the S.A.T. for Women's Scores" The New York Times, June 19.

Rosser, Phyllis. 1989. The SAT Gender Gap: Identifying the Causes. Washington, D.C.: Center for Women Policy Studies.

Rosser, Phyllis. 1987a. Sex Bias in College Admissions Tests: Why Women Lose Out. Cambridge, Massachusetts: National Center for Fair and Open Testing.

Rosser, Phyllis. 1987b. Congressional Testimony before the House Subcommittee on Civil and Constitutional Rights, Washington, D.C., April 23.

Schmitt, Alicia, and Carole Bleistein. 1986. "Factors Affecting Differential Item Functioning for Black Examinees on Scholastic Aptitude Test Analogy Items." Princeton: Educational Testing Service.

Selkow, Paula. 1984. Assessing Sex Bias in Testing: A Review of the Issues and Evaluations of 74 Psychological and Educational Tests. Westport: Greenwood Press.

Sternberg, Robert J. 1977. Intelligence, Information Processing, and Analogical Reasoning: the Componential Analysis of Human Abilities. Hillsdale: Lawrence Erlbaum Associates.

Tittle, Carol Kehr. 1978. Sex Bias in Testing: A Review with Policy Recommendations. San Francisco: Women's Educational Equity Communications Network, Far West Laboratory.

Tittle, Carol K., K. McCarthy, and J.F. Streckler. 1974. "Women and Educational Testing." Princeton: Educational Testing Service.

APPENDICES

APPENDIX A
PROFILES OF STUDENTS PARTICIPATING IN STUDY

	5 WHITE FEMALES					5 BLACK FEMALES				
Identifying number	1	2	3	4	5	6	7	8	9	10
Gender	F	F	F	F	F	F	F	F	F	F
Ethnic Group	W	W	W	W	W	B	B	B	B	B
Number correct out Of 8	5	3	4	1	3	1	4	1	0	3
Cumulative GPA	3.4	3.5	3.8	2.5	3.6	3	3	2.6	2.8	3
English GPA	3.6	4	3.8	3	3.7	3.2	3	2.5	2.9	3.7
Previously coached	1	0	0	0	0	0	0	0	0	1
Inquired about scoring	0	1	0	0	0	0	1	0	0	0
Negative ability comments	0	1	0	1	0	0	0	1	1	0

	5 WHITE MALES					5 BLACK MALES				
Identifying number	11	12	13	14	15	16	17	18	19	20
Gender	M	M	M	M	M	M	M	M	M	M
Ethnic Group	W	W	W	W	W	B	B	B	B	B
Number correct out of 8	1	2	5	6	1	4	6	3	1	2
Cumulative GPA	3	2.8	3.3	4	3.1	2.5	3.8	3	2.8	2.9
English GPA	3	3.4	3.6	4	3	3.5	3.6	2.6	3	2.7
Previously coached	0	0	1	1	0	0	0	0	0	0
Inquired about scoring	1	0	0	1	0	1	1	0	1	1
Negative ability comments	0	1	0	0	0	0	0	0	0	0

APPENDIX B - QUESTIONNAIRE GIVEN TO EACH STUDENT PARTICIPATING IN STUDY

PILOT STUDY - SAT VERBAL

PILOT STUDY - SAT VERBAL QUESTIONS

Thank you for taking time to participate in this study. Please complete the following information:

Name_____ M F Date _____

Age:_____ Grade: _____ GPA:_____ Ethnic group:_____

Have you taken the PSAT? Yes No If yes, please note your scores:

Verbal PSAT _____ Math PSAT _____

Have you taken the SAT? Yes No If yes, please note your scores:

Verbal SAT _____ Math SAT _____

Do you plan to take the SAT within the next six months? Yes No

Have you ever attended an SAT coaching class? Yes No

Have you ever read a book or used a computer program on how to improve your performance on the SAT? Yes No

What is your ESTIMATED high school GPA in English/Language Arts Literature/Composition classes? _____

READ THE DIRECTIONS BELOW AND COMPLETE THE SAMPLE ITEM:

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five lettered words or sets of words. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

Example:
Although its publicity has been ----, the film itself is intelligent, well-acted, handsomely produced, and altogether ----.

(A) tasteless..respectable (B) extensive..moderate
(C) sophisticated..amateur (D) risque..crude
(E) perfect..spectacular

(answer is A)

TELL ME WHEN YOU HAVE COMPLETED THE DIRECTIONS AND SAMPLE ITEM.

NOW ANSWER THE THREE ITEMS BELOW. TELL ME AS YOU FINISH EACH ONE.

4-23. The ruling by the Supreme Court that the appropriation by a state government of Cherokee lands was unconstitutional did not ---- such seizures because President Jackson failed to ---- it.

(A) legitimize..ratify (B) change..explicate
(C) justify..challenge (D) end..commute
(E) halt..implement

4-24. When people habitually read or study with music playing in the background, they unfortunately cultivate "passive listening," which is the very ---- of listening, a process that is essentially active.

(A) presage (B) antithesis (C) circumstance
(D) pinnacle (E) illumination

4-25. Although the undefeated visitors ----triumphed over their underdog opponents, the game was hardly the ---- sportswriters had predicted.

(A) fortunately..upset (B) unexpectedly..classic
(C) finally..rout (D) easily..stalemate
(E) utterly..mismatch

PLEASE RATE EACH ITEM AS TO DIFFICULTY:

4-23. Very Difficult Difficult Not Sure Easy Very Easy
Would you have left this item out on the SAT? yes no

4-24. Very Difficult Difficult Not Sure Easy Very Easy
Would you have left this item out on the SAT? yes no

4-25. Very Difficult Difficult Not Sure Easy Very Easy
Would you have left this item out on the SAT? yes no

READ THE DIRECTIONS BELOW AND COMPLETE THE SAMPLE ITEM:

Each question below consists of a related pair of words or phrases, followed by five lettered pairs of words or phrases. Select the lettered pair that best expresses a relationship similar to that expressed in the original pair.

Example:
YAWN : BOREDOM :: (A) dream : sleep
(B) anger : madness (C) smile : amusement
(D) face : expresssion (E) impatience : rebellion

(answer is C)

TELL ME WHEN YOU COMPLETE THE DIRECTIONS AND SAMPLE ITEM.

NOW ANSWER THE THREE ITEMS BELOW. TELL ME AS YOU COMPLETE EACH ONE.

1-21. OVERWHELM : FORCE ::
(A) conquer : victory
(B) relieve : pressure
(C) entrap : deceit
(D) surrender : strength
(E) bully : submission

1-23. DESSICATE : MOISTURE ::
(A) displace : volume
(B) engender : originality
(C) implicate : logic
(D) dispirit : morale
(E) inundate : fluid

4-41. DIVIDENDS : STOCKHOLDERS ::
(A) investments : corporations
(B) purchases : customers
(C) royalties : authors
(D) taxes : workers
(E) mortgages : homeowners

PLEASE RATE EACH ITEM AS TO DIFFICULTY:

1-21.	Very Difficult	Difficult	Not Sure	Easy	Very Easy
Would you have left this item out on the SAT?				yes	no
1-23.	Very Difficult	Difficult	Not Sure	Easy	Very Easy
Would you have left this item out on the SAT?				yes	no
4-41.	Very Difficult	Difficult	Not Sure	Easy	Very Easy
Would you have left this item out on the SAT?				yes	no

READ THE DIRECTIONS BELOW AND COMPLETE THE SAMPLE ITEM:

Each question below consists of a word in capital letters, followed by five lettered words or phrases. Choose the word or phrase that is most nearly opposite in meaning to the word in capital letters. Since some of the questions require you to distinguish fine shades of meaning, consider all the choices before deciding which is best.

Example:
GOOD: (A) sour (B) bad (C) red
(D) hot (E) ugly

(answer is B)

TELL ME WHEN YOU COMPLETE THE DIRECTIONS AND SAMPLE ITEM.

NOW ANSWER THE TWO ITEMS BELOW. TELL ME AS YOU COMPLETE EACH ONE.

1-7. PREVALENT: (A) limited (B) confused
(C) inexact (D) important (E) healthy

1-8. HAPHAZARD: (A) resistant (B) original
(C) secure (D) considered (E) confidential

PLEASE RATE EACH ITEM AS TO DIFFICULTY:

1-7.	Very Difficult	Difficult	Not Sure	Easy	Very Easy
Would you have left this item out on the SAT?				yes	no

1-8.	Very Difficult	Difficult	Not Sure	Easy	Very Easy
Would you have left this item out on the SAT?				yes	no

APPENDIX D

SCORING SHEET FOR PILOT STUDY - SAT VERBAL

Name_____M F Ethnic_____

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five lettered words or sets of words. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

4-23. The ruling by the Supreme Court that the appropriation by a state government of Cherokee lands was unconstitutional did not ---- such seizures because President Jackson failed to ---- it.

(A) legitimize..ratify (B) change..explicate
(C) justify..challenge (D) end..commute
(E) halt..implement

ANSWER CHOSEN:_____ (correct answer: E) TIME_____

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
  -- tried placing pair choices into sentence
  -- eliminated pairs: (___) (___) (___) (___)
    -- because they didn't "make sense"
    -- because they didn't "sound right"
    -- because didn't know meaning of words
  -- narrowed choice to two pairs (between ___ and ___)

--selected answer
  -- because it "made the most sense"
  -- because it "sounded right"
  -- after narrowing choices to two
  -- after eliminating all other choices

--selected correct answer earlier, then changed mind
  reason: _____

--comprehended main idea of stem sentence
--knew meaning of ___ out of 28 words in stem sentence
  did not know _____)
--looked for key words in stem sentence to use as clues
  key words _____

--had studied this historical event in school so knew about it

--knew meaning of ___ out of the ten words in the pairs
  --legitimize.... --ratify
  --change.... --explicate
  --justify.... --challenge
  --end.... --commute
  --halt.... --implement

--analyzed unknown words to determine meaning
  -- broke words into pieces
  -- compared to similar known words
  -- tried to remember context in which seen it used before

4-24. When people habitually read or study with music playing in the background, they unfortunately cultivate "passive listening," which is the very ---- of listening, a process that is essentially active.

(A) presage (B) antithesis (C) circumstance
(D) pinnacle (E) illumination

ANSWER CHOSEN_____(correct answer: B) TIME_____

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
    -- tried placing each choice word into sentence
    -- eliminated choices: (___) (___) (___) (___)
        -- because they didn't "make sense"
        -- because they didn't "sound right"
        -- because didn't know meaning of words
    -- narrowed choices to two words (between ___ and ___)

--selected answer
    -- because it "made the most sense"
    -- because it "sounded right"
    -- after narrowing choices to two
    -- after eliminating all other choices

--selected correct answer earlier, then changed mind
    reason: _____

--comprehended main idea of stem sentence
--knew meaning of ___ out of 29 words in stem sentence
    did not know _____)
--looked for key words in stem sentence to use as clues
    key words _____

--had studied this concept in school so knew about it

--knew meaning of ___ out of the five choice words

  -- presage
  -- antithesis
  -- circumstance
  -- pinnacle
  -- illumination

--analyzed unknown words to determine meaning
    -- broke words into pieces
    -- compared to similar known words
    -- tried to remember context in which seen it used before

4-25. Although the undefeated visitors ----triumphed over their underdog opponents, the game was hardly the ---- sportswriters had predicted.

(A) fortunately..upset (B) unexpectedly..classic
(C) finally..rout (D) easily..stalemate
(E) utterly..mismatch

ANSWER CHOSEN_____(correct answer:C) TIME_____

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
-- tried placing pair choices into sentence
-- eliminated pairs: (___) (___) (___) (___)
-- because they didn't "make sense"
-- because they didn't "sound right"
-- because didn't know meaning of words
-- narrowed choice to two pairs (between ___ and ___)

--selected answer
-- because it "made the most sense"
-- because it "sounded right"
-- after narrowing choices to two
-- after eliminating all other choices

--selected correct answer earlier, then changed mind
reason: _____

--comprehended main idea of stem sentence
--knew meaning of ___ out of 17 words in stem sentence
did not know _____)
--looked for key words in stem sentence to use as clues
key words _____

--familiar with sports world so understood situation

--knew meaning of ___ out of the ten words in the pairs

-- fortunately -- upset
-- unexpectedly -- classic
-- finally -- rout
-- easily -- stalemate
-- utterly -- mismatch

--analyzed unknown words to determine meaning
-- broke words into pieces
-- compared to similar known words
-- tried to remember context in which seen it used before

Each question below consists of a related pair of words or phrases, followed by five lettered pairs of words or phrases. Select the lettered pair that best expresses a relationship similar to that expressed in the original pair.

1-21. OVERWHELM : FORCE ::
(A) conquer : victory
(B) relieve : pressure
(C) entrap : deceit
(D) surrender : strength
(E) bully : submission

ANSWER CHOSEN_____(correct answer:C) TIME_____

--understood idea of analogies

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
-- eliminated pairs: (___) (___) (___) (___)
-- because relationship was not like original pair
-- because couldn't tell what relationship was
-- because didn't know meaning of words in pair
-- narrowed choice to two pairs (between ___ and ___)

--selected answer
-- because relationship seemed similar to original pair
-- because meaning of words was similar to meaning of words in original pair
-- after narrowing choices to two
-- after eliminating all other choices

--selected correct answer earlier, then changed mind
reason: _____

knew meaning of ___ out of 2 words in original pair
--understood relationship between words in original pair
--analyzed words in original pair to determine relationship
--made words into sentence "_____
_____"
--described type of relationship "_____
_____"

knew meaning of ___ out of 10 words in choice pairs
--understood relationship between words in ___ out of the five choice pairs
--analyzed choice pairs to determine relationships
--made words into sentence "_____
_____"
--described type of relationship "120
_____"

1-23. DESSICATE : MOISTURE ::
(A) displace : volume
(B) engender : originality
(C) implicate : logic
(D) dispirit : morale
(E) inundate : fluid

ANSWER CHOSEN___(correct answer:D) TIME_____

--understood idea of analogies

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
-- eliminated pairs: (___) (___) (___) (___)
-- because relationship was not like original pair
-- because couldn't tell what relationship was
-- because didn't know meaning of words in pair
-- narrowed choice to two pairs (between ___ and ___)

--selected answer
-- because relationship seemed similar to original pair
-- because meaning of words was similar to meaning of words in original pair
-- after narrowing choices to two
-- after eliminating all other choices

--selected correct answer earlier, then changed mind
reason: _____

knew meaning of ___ out of 2 words in original pair
--understood relationship between words in original pair
--analyzed words in original pair to determine relationship
--made words into sentence "_____

_____"

--described type of relationship "_____

_____"

knew meaning of ___ out of 10 words in choice pairs
--understood relationship between words in ___ out of the five choice pairs
--analyzed choice pairs to determine relationships
--made words into sentence "_____

_____"

--described type of relationship "_____

_____"

4-41. DIVIDENDS : STOCKHOLDERS ::
(A) investments : corporations
(B) purchases : customers
(C) royalties : authors
(D) taxes : workers
(E) mortgages : homeowners

ANSWER CHOSEN ----(correct answer:C) TIME_____

--understood idea of analogies

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
-- eliminated pairs: (___) (___) (___) (___)
-- because relationship was not like original pair
-- because couldn't tell what relationship was
-- because didn't know meaning of words in pair
-- narrowed choice to two pairs (between ___ and ___)

--selected answer
-- because relationship seemed similar to original pair
-- because meaning of words was similar to meaning of words in original pair
-- after narrowing choices to two
-- after eliminating all other choices

--selected correct answer earlier, then changed mind
reason: _____

knew meaning of ___ out of 2 words in original pair
--understood relationship between words in original pair
--analyzed words in original pair to determine relationship
--made words into sentence "_____

_____"

--described type of relationship "_____

_____"

knew meaning of ___ out of 10 words in choice pairs
--understood relationship between words in ___ out of the five choice pairs
--analyzed choice pairs to determine relationships
--made words into sentence "_____

_____"

--described type of relationship "_____

_____"

Each question below consists of a word in capital letters, followed by five lettered words or phrases. Choose the word or phrase that is most nearly opposite in meaning to the word in capital letters. Since some of the questions require you to distinguish fine shades of meaning, consider all the choices before deciding which is best.

1-7. PREVALENT: (A) limited (B) confused
(C) inexact (D) important (E) healthy

ANSWER CHOSEN_____(correct answer:A) TIME_____

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
-- eliminated choices: (___) (___) (___) (___)
-- because they didn't "make sense as an opposite"
-- because they didn't "sound right as an opposite"
-- because didn't know their meaning
-- narrowed choice to two (between ___ and ___)

--selected answer
-- because it "made the most sense as an opposite"
-- because it "sounded right as an opposite"
-- because it was the "closest to being opposite"
-- after narrowing choices to two
-- after eliminating all others

--selected correct answer earlier, then changed mind
reason: _____

--understood meaning of stem word

"it means _____"

--analyzed stem word to determine meaning
-- broke word into pieces
-- compared to similar known words
-- tried to remember context in which seen/heard it used before

--understood meaning of ___ out of the five choice words

did not know _____

--analyzed unknown words to determine meaning
-- broke words into pieces
-- compared to similar known words
-- tried to remember context in which seen/heard it used before

1-8. HAPHAZARD: (A) resistant (B) original
(C) secure (D) considered (E) confidential

ANSWER CHOSEN _____(correct answer:D) TIME_____

--answer came quickly to mind - "just knew it"
--answer was a wild guess
--answer arrived at by a process
-- eliminated choices: (___) (___) (___) (___)
-- because they didn't "make sense as an opposite"
-- because they didn't "sound right as an opposite"
-- because didn't know their meaning
-- narrowed choice to two (between ___ and ___)

--selected answer
-- because it "made the most sense as an opposite"
-- because it "sounded right as an opposite"
-- because it was the "closest to being opposite"
-- after narrowing choices to two
-- after eliminating all others

--selected correct answer earlier, then changed mind
reason: _____

--understood meaning of stem word

"it means _____"

--analyzed stem word to determine meaning
-- broke word into pieces
-- compared to similar known words
-- tried to remember context in which seen/heard it used before

--understood meaning of ___ out of the five choice words

did not know _____

--analyzed unknown words to determine meaning
-- broke words into pieces
-- compared to similar known words
-- tried to remember context in which seen/heard it used before

Appendix E

Five Probelms Used in Math Study

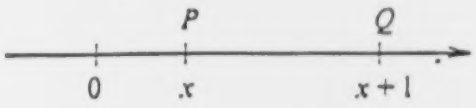
The five problems included in this study were taken from the November 1987 SAT test (code form 7H). The test items are: Problem 1 - section 2, number 16; Problem 2 - section 2, number 17; Problem 3 - section 5, number 6; Problem 4 - section 5, number 29; and Problem 5 - section 5, number 31.

Appendix E

Name: _____

(SECTION 2 – QUESTION 16)

1)



On the number line above, which of the following is the coordinate of the midpoint of segment PQ?.

(A) $\frac{x}{2}$

(B) $\frac{x+1}{2}$

(C) $x + \frac{1}{2}$

(D) $2(x + 1)$

(E) $\frac{x(x+1)}{2}$

Name: _____

2) Oatmeal Recipe

(SECTION 2 - QUESTION 17)

Water: $\frac{3}{4}$ cup
Salt: $\frac{1}{4}$ teaspoon
Oats: $\frac{1}{3}$ cup

If the least possible multiple of the recipe above is prepared so that a whole number of cups of both water and oats are used, how many teaspoons of salt would be required?

(A) $\frac{1}{2}$

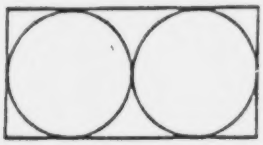
(B) $\frac{3}{4}$

(C) 1

(D) $2\frac{1}{4}$

(E) 3

Name: _____

3) 

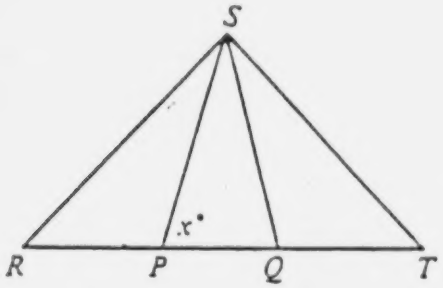
(SECTION 5 - QUESTION 6)

The rectangle above contains two circles, tangent to each other and each tangent to three sides of the rectangle. Which of the following pairs of numbers CANNOT be the length and width, respectively, of the rectangle?

(A) 2, 1
(B) 12, 6
(C) 16, 10
(D) 22, 11
(E) 32, 16

Name: _____

4)



(SECTION 5 - QUESTION 29)

Note: Figure not drawn to scale.

In the figure above, $\triangle RST$ is a right triangle. $RS = ST$ and right angle RST has been divided into three equal angles. What is the value of x?

(A) 65
(B) 70
(C) 75
(D) 80
(E) 85

Name: _____

5) If a rectangular cake, 9 inches by 13 inches by 2 inches, is cut into x equal rectangular pieces, 3 inches by $3\frac{1}{4}$ inches by 2 inches, and no cake is left over, then x =

(SECTION 5 - QUESTION 31)

(A) 9
(B) 12
(C) 13
(D) 15
(E) 22

Appendix F

Process-Sequence Coding System for SAT Interviews[1]

Processes

R	Reads the problem
S	Separates or summarizes the data
M	Marks diagram
D	Draws diagram
Q	Rephrases question
P_F	Suggests plan for final goal
P_I	Suggests plan for intermediate goal
P_R	States rules/theorem/facts to use
P_O	States operation to use
V	Identifies variable
P_E	Sets up equation
A	Uses algorithm
SB	Solves by substitution
L_I	Uses induction
L_D	Uses deduction
T_R	Uses random trial and error
T_S	Uses trial and error - successive approximations
G	Guesses
C	Checks solution
C_A	Introduces alternate procedure - varies the problem
X	Has forgotten or does not know how to do the problem
N	Not classifiable

Outcomes

1	Abandons process
2	Impasse
3	Correct final results
4	Correct intermediate result
5	Incorrect result

Punctuation

,	Inserted between successive processes
—	Hesitation of approximately 30 seconds
/	Stops without solution
-	Stops with solution (correct or incorrect)
( )	Scope of a process

Errors - written over process symbols

:—	Structural error in process
:	Executive error in process
**	Previous error was corrected
_	Difficulty with process (underline)

---

[1]This list is a combination of Lucas' (1972), Kantowski's (1974), and Blake's (1976) Process-Sequence Coding System. From - McClintock, E. (1984). Heuristics processes as task variables. In G. Goldin & E. McClintock (Eds.) Task variables in mathematical problem solving (pp.171-234). Philadelphia, PA: Franklin Institute Press.

Student's Name: _____ Age: _____
Grade: _____

Math PSAT: _____ Verbal PSAT: _____ GPA: _____

Permission to record: Yes/No Tape count position _____

Start Time: _____

---

1) Observations - to be completed while student is working on problem 1 (see Coding sheet):

Start Time: _____ Time Finished: _____

---

2) Observations - to be completed while student is working on problem 2 (see Coding sheet):

Start Time: _____ Time Finished: _____

---

3) Observations - to be completed while student is working on problem 3 (see Coding sheet):

Start Time: _____ Time Finished: _____

Student ID: _____

4) Observations - to be completed while student is working on problem 4 (see Coding sheet):

Start Time: _____ Time Finished: _____

_____

_____

_____

_____

_____

_____

5) Observations - to be completed while student is working on problem 5 (see Coding sheet):

Start Time: _____ Time Finished: _____

_____

_____

_____

_____

_____

Student ID: _____

Interview (to be completed after student is finished working on problems):

Question #1

How would you rate this problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

How would you rate this as an SAT problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

Please explain to me step by step how you went about solving this problem. (Follow-up(s) - Please explain to me how you tried to solve this problem. Please explain how you arrived at the answer you circled.)

Student ID: _____

Question #2

How would you rate this problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

How would you rate this as an SAT problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

Please explain to me step by step how you went about solving this problem. (Follow-up(s) - Please explain to me how you tried to solve this problem. Please explain how you arrived at the answer you circled.)

Student ID: _____

Question #3

How would you rate this problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

How would you rate this as an SAT problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

Please explain to me step by step how you went about solving this problem. (Follow-up(s) - Please explain to me how you tried to solve this problem. Please explain how you arrived at the answer you circled.)

Student ID: _____

Question #4

How would you rate this problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

How would you rate this as an SAT problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

Please explain to me step by step how you went about solving this problem. (Follow-up(s) - Please explain to me how you tried to solve this problem. Please explain how you arrived at the answer you circled.)

Student ID: _____

Question #5

How would you rate this problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

How would you rate this as an SAT problem (choose one):
Very difficult Difficult Not Sure Easy Very Easy

Please explain to me step by step how you went about solving this problem. (Follow-up(s) - Please explain to me how you tried to solve this problem. Please explain how you arrived at the answer you circled.)

Time finished: _____ Tape Count: _____

Please respond to each of the following statements:

1) The verbal section of the PSAT was (circle one):
Extremely Difficult for me Difficult for me Not Sure /Don't Know Easy for me Extremely Easy for me

2) The math section of the PSAT was (circle one):
Extremely Difficult for me Difficult for me Not Sure /Don't Know Easy for me Extremely Easy for me

For Release Tuesday, September 20, 1988

(from News From the College Board)

College-Bound Seniors

SAT* Score Averages, 1967-1988

Year	Verbal			Mathematical		
	Men	Women	Total	Men	Women	Total
1967	*463*	*468*	*466*	*514*	*467*	*492*
1968	*464*	*466*	*466*	*512*	*470*	*492*
1969	*459*	*466*	*463*	*513*	*470*	*493*
1970	*459*	*461*	*460*	*509*	*465*	*488*
1971	*454*	*457*	*455*	*507*	*466*	*488*
1972	454	452	453	505	461	484
1973	446	443	445	502	460	481
1974	447	442	444	501	459	480
1975	437	431	434	495	449	472
1976	433	430	431	497	446	472
1977	431	427	429	497	445	470
1978	433	425	429	494	444	468
1979	431	423	427	493	443	467
1980	428	420	424	491	443	466
1981	430	418	424	492	443	466
1982	431	421	426	493	443	467
1983	430	420	425	493	445	468
1984	433	420	426	495	449	471
1985	437	425	431	499	452	475
1986	437	426	431	501	451	475
1987	435	425	430	500	453	476
1988	435	422	428	498	455	476

The averages for 1967 through 1971 are estimates. College-Bound Seniors reports were not prepared in those years.

SAT SCORE AVERAGES FOR THE CLASS OF 1988

from 1988 National Sex/Ethnic Profiles by College Entrance Examination Board

GROUP (NUMBER OF TEST TAKERS)	VERBAL	MATH	COMBINED	DIFFERENCE
White Males (388,478)	452	513	965	
White Females (424,638)	439	468	907	- 58
Asian/American Males (32,917)	413	543	956	
Asian/American Females (31,185)	404	499	903	- 53
Latin American Males (9,398)	395	459	854	
Latin American Females (10,815)	380	411	791	- 63
Native American Males (5,906)	398	454	852	
Native American Females (6,424)	387	418	805	- 47
Mexican Males (10,544)	391	449	840	
Mexican Females (12,178)	373	410	783	- 57
Puerto Rican Males (5,051)	364	424	788	
Puerto Rican Females (6,446)	348	384	732	- 56
Black Males (40,099)	358	398	756	
Black Females (57,384)	350	374	724	- 32
Other Males (6,634)	419	489	908	
Other Females (7,460)	401	434	835	- 73
ALL MALES (580,915)	435	498	933	
ALL FEMALES (632,304)	422	455	877	- 56

APPENDIX J

1987-88 ACT Composite Scores

White males	21.0
Asian American males	20.8
White females	19.6
Asian American females	19.4
Puerto Rican males	18.1
Puerto Rican females	16.6
Native American males	16.6
Mexican American males	16.3
Native American females	15.1
Mexican American females	14.8
African American males	14.1
African American females	13.4

(Source: *College Student Profiles: Norms for the ACT Assessment*, 1987)

APPENDIX K

Females' Average Scores Are Lower Than Males' At Each Income Level

1988 National Sex/Ethnic Profiles
College Entrance Examination Board

SAT Scores By Family Income

1988 Profiles for Males	Number of SAT Takers	Percent	SAT-V Mean	SAT-M Mean	Combined Score
Income					
Less than 10,000	19,442	4	372	450	822
10,000-20,000	52,936	11	400	465	865
20,000-30,000	78,175	17	422	483	905
30,000-40,000	93,073	20	434	494	928
40,000-50,000	67,094	14	446	508	954
50,000-60,000	49,880	11	455	518	973
60,000-70,000	30,212	6	461	524	985
70,000 or more	74,494	16	474	542	1016
No response	78,759				

1988 Profiles for Females	Number of SAT Takers	Percent	SAT-V Mean	SAT-M Mean	Combined Score
Income					
Less than 10,000	28,835	6	356	397	753
10,000-20,000	69,798	14	387	421	808
20,000-30,000	88,877	18	410	440	850
30,000-40,000	100,105	20	423	453	876
40,000-50,000	68,885	14	436	467	903
50,000-60,000	50,240	10	444	477	921
60,000-70,000	30,017	6	451	485	936
70,000 or more	69,614	14	463	502	965
No response	83,928				

FEMALES' AVERAGE SCORES ARE LOWER THAN MALES AT EACH EDUCATIONAL LEVEL

1988 National Sex/Ethnic Profiles
College Entrance Examination Board

SAT Scores by Family Educational Level

1988 Profiles for Males

Highest Level of Parental Education	Number of SAT Takers	Percent	SAT-V Mean	SAT-M Mean	Combined Score
No High School Diploma	18,575	4	355	438	793
High School Diploma	174,450	35	408	469	877
Associate Degree	33,983	7	419	478	897
Bachelor's Degree	141,250	29	452	517	969
Graduate Degree	125,490	25	482	546	1028

1988 Profiles for Females

Highest Level of Parental Education	Number of SAT Takers	Percent	SAT-V Mean	SAT-M Mean	Combined Score
No High School Diploma	25,398	5	341	389	730
High School Diploma	212,647	39	397	428	825
Associate Degree	38,224	7	409	439	848
Bachelor's Degree	144,555	26	441	475	916
Graduate Degree	127,813	23	470	502	972

APPENDIX L - AVERAGE SAT SCORES FOR MALES AND FEMALES ON NOVEMBER 1987 SAT BY GRADE POINT AVERAGE

(100,000 SAMPLE)

Average SAT Scores for Females and Males in Each GPA Category

GPA	Female [Number of Students]	Male	Difference
Verbal Averages by GPA			
A+ to A	514 (7,492)	537 (5,406)	- 23
A- to B+	457 (17,033)	477 (12,033)	- 20
B to B-	404 (19,387)	425 (15,895)	- 21
C+ to F	363 (7,330)	382 (8,408)	- 19
Math Averages By GPA			
A+ to A	564	624	- 60
A- to B+	495	554	- 59
B to B-	430	481	- 51
C+ to F	378	421	- 43

Source: The SAT Gender Gap: Identifying the Causes
April 1989

November 1987 SAT Score Averages by Sex and Race (100,000 Sample)

GROUP	VERBAL	MATH	COMBINED SCORE	FEMALE DIFF.
White Males	457.01	516.61	973.62	
33,620				
White Females	443.51	473.23	916.74	- 56.88
40,846				
Asian/Amer. Males	416.48	542.81	959.29	
2,694				
Asian/Amer. Females	408.54	502.87	911.41	- 47.88
2,724				
Native Amer. Males	410.28	470.89	881.17	
429				
Native Amer. Females	395.67	418.44	814.11	- 67.06
601				
Hispanic Males	397.67	457.10	854.77	
1,791				
Hispanic Females	379.53	405.88	785.41	- 69.36
2,373				
Black Males	380.50	414.12	794.62	
2,829				
Black Females	371.07	387.53	758.60	- 36.02
4,441				
Other Males	428.58	484.86	913.44	
4,028				
Other Females	416.54	447.29	863.83	- 49.61
3,621				